U.S. Army Research Institute
for the Behavioral and Social Sciences

Research Report 1854

Behaviorally Anchored Rating Scales for the Assessment of Tactical Thinking Mental Models

Jennifer K. Phillips, Jennifer Shafer,
Karol G. Ross and Donald A. Cox
Klein Associates

Scott B. Shadrick
U.S. Army Research Institute

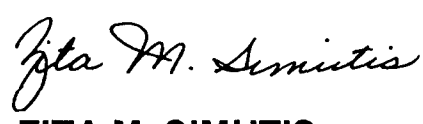
June 2006

U.S. Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE SAMS
Technical Director



ZITA M. SIMUTIS
Director

Research accomplished under contract
for the Department of the Army

Technical review by

William R. Sanders, U.S. Army Research Institute
Gregory Goodwin, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Research Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Research Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Research Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
June 2006	Final	March 2004 to December 2005

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Behaviorally Anchored Rating Scales for the Assessment of Tactical Thinking Mental Models	W74V8H-04-C-0018
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Jennifer K. Phillips, Jennifer Shafer, Karol G. Ross, Donald A. Cox (Klein Associates), and Scott B. Shadrick (U.S. Army Research Institute for the Behavioral and Social Sciences)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 211
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Klein Associates 1750 Commerce Center Blvd N. Fairborn, OH 45324-6362	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-IK 121 Morande Street Fort Knox, KY 40121-4141	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Research Report 1854

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Contracting Officer's Representative (COR) and Subject Matter POC: Scott B. Shadrick

14. ABSTRACT *(Maximum 200 words)*
Report developed under a Broad Agency Announcement (BAA) # W74V8H-04-C-0018. An ongoing need exists in the Army to enhance combat leaders' tactical thinking skills. In conjunction, measurement techniques must be developed to assess tactical thinking skills. This report documents an effort to develop a standardized and reliable assessment tool for purposes of evaluating training applications, diagnosing individuals' levels of cognitive proficiency, and examining the impact of advanced battle command technologies on user cognition. Four Tactical Thinking Behaviorally Anchored Rating Scales (T-BARS) were developed. They enable researchers to measure cognitive proficiency along critical dimensions of tactical thinking by coding behaviors that are observable in the context of training sessions, exercises, or experiments. Themes of tactical thinking identified in the Think Like A Commander program formed the basis of T-BARS. The Dreyfus & Dreyfus (1986) stage model of cognitive skill acquisition guided construct development for five levels of tactical thinking proficiency within each scale. Interviews were conducted with Army officers to elicit patterns of thinking and behaviors in tactical exercises. Interview data informed the behavioral descriptors generated to populate the levels of cognitive performance within the T-BARS. Scale development occurred iteratively with interrater reliability testing. The finalized T-BARS assessment tool is accompanied by a user guide to support its application.

15. SUBJECT TERMS
Tactical thinking; cognitive skills assessment; assessment; cognitive skills evaluation; behaviorally anchored rating scales; scenario-based training assessment; cognitive training; decision training

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	62	Ellen Kinzer Technical Publication Specialist 703-602-8047

Research Report 1854

Behaviorally Anchored Rating Scales for the Assessment of Tactical Thinking Mental Models

Jennifer K. Phillips, Jennifer Shafer,
Karol G. Ross and Donald A. Cox
Klein Associates

Scott B. Shadrick
U.S. Army Research Institute

Armored Forces Research Unit
Barbara A. Black, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, VA 22202-3926

June 2006

Army Project Number
622785A790

Personnel Performance
and Training Technology

BEHAVIORALLY ANCHORED RATING SCALES FOR THE ASSESSMENT OF TACTICAL THINKING MENTAL MODELS

EXECUTIVE SUMMARY

Research Requirement:

An ongoing need exists in the Army to enhance combat leaders' tactical thinking skills. By "thinking skills" we refer to the higher-order cognitive functions such as decision making, sense making, and the underlying cognitive processes that support those functions. To improve cognitive task performance, Soldiers and leaders often engage in scenario-based training sessions that allow deliberate decision making practice in context-rich environments. A critical aspect of deliberate training for tactical thinking skills that requires further development is assessment. How do we know that thinking skills are improving across experiences and over time? Current assessment techniques used in military training rely on either objective measures that do not reflect the underlying cognitive skills, or subjective domain experts' judgments that are difficult to standardize and often difficult to obtain. There is a need for an assessment tool that will allow us to measure the development of thinking skills more objectively and reliably. The research effort documented in this report addresses this need by developing a behaviorally anchored rating scale for tactical thinking mental models.

Procedure:

A Tactical Thinking Behaviorally Anchored Rating Scale (T-BARS) was generated and interrater reliability was established. Eight themes of tactical thinking identified in the Think Like A Commander program of research and training formed the basis of the scales. The Dreyfus & Dreyfus (1986) stage model of cognitive skill acquisition guided construct development for five levels of tactical thinking proficiency within each scale. Interviews were conducted with Army officers with a range of operational experience to elicit patterns of thinking and behaviors within a set of tactical exercises. Interview data were utilized to generate behavioral indicators to populate the five levels of cognitive performance within the T-BARS. Scale development occurred iteratively with interrater reliability testing, as results of the testing informed the next version of the scales. Once T-BARS were finalized, a User Guide was produced to support application of the assessment tool for training evaluation and other purposes.

Findings:

The finalized T-BARS tool contains four scales representing tactical thinking mental models: Know and Use All Assets Available; Consider the Mission and Higher's Intent; Model a Thinking Enemy; and Consider Effects of Terrain. Five levels of cognitive performance are accounted for within each scale: novice; advanced beginner; competent; proficient; and expert. A set of behavioral descriptors are associated with each of the five levels of performance, enabling linkages to be made between actions that are observed during training sessions or exercises and the performer's cognitive proficiency. Results of the interrater reliability testing show that the ratings are consistent and hold together to measure common dimensions. Rater

consensus when coding for tactical thinking mental models was high. Consensus when coding for levels within a particular mental model scale was high when single category differentials between judges were allowed.

Utilization and Dissemination of Findings:

T-BARS is intended for primary use by researchers who are versed in naturalistic cognition and familiar with the military domain. It can be applied to assess verbal protocol data, written measures of performance (such as courses of action and orders), or performance during exercise observations. The value of T-BARS is that it provides a standard technique for measuring an individual's cognitive proficiency. The results of a T-BARS assessment can be used to diagnose an individual's tactical skills to determine an appropriate track of training; measure the impact of a training intervention on cognitive performance to assess the effectiveness of the intervention; or measure the impact of a new technology on cognitive performance to assess the value of the technology.

BEHAVIORALLY ANCHORED RATING SCALES FOR THE ASSESSMENT OF TACTICAL THINKING MENTAL MODELS

CONTENTS

Page

List of Tables

List of Figures

BEHAVIORALLY ANCHORED RATING SCALES FOR THE ASSESSMENT OF TACTICAL THINKING MENTAL MODELS

Introduction

The cultivation of cognitive skills is central to developing expertise in complex, ill-structured domains such as military tactics. Within these domains performance depends on declarative knowledge in the form of facts; procedural knowledge with regard to employing weapon systems or implementing specific techniques; and tacit or implicit knowledge, which refers to the higher order task of assessing the operational environment and deciding *how, when,* or *where* to implement tactics to achieve the desired result. While declarative and procedural knowledge are relatively amenable to measurement due to their objective nature, the cognitive skills that propel effective decision making and assessment are challenging to quantify and measure. Cognition and thought cannot be seen by an observer; only the outcomes of those processes are observable. However, given the criticality of cognitive skills in the performance of tactical and other complex tasks, it is necessary to develop a means of measurement to inform human support activities such as training or technology development.

Assessment in training applications is largely focused on declarative and procedural knowledge. However, the training community is also creating interventions that target thinking skills—higher order cognitive functions such as decision making, sense making, and the underlying cognitive processes such as problem detection that support those functions. In complex, ill-structured domains such as tactical thinking, medical diagnosis and treatment, and law enforcement, it is not enough to rely on rote procedures and factual knowledge. Operators require declarative and procedural knowledge as foundations, but different situations in these domains are likely to require the application of varying patterns of principles, even in cases of seemingly similar problems or goals. No standard solutions can be employed with regularity. Such domains require professionals to exercise a great deal of judgment to flexibly apply their knowledge. Well-developed thinking skills are critical for high levels of performance. These skills, too, need to be trained, and for effective training we must be able to assess them.

Beyond assessment of training interventions, there is a pronounced need to evaluate the impact of advanced technologies on human cognitive performance. The military spends millions of dollars on battle command systems intended to support tactical decision making through visualization technologies, planning software, and other tools. The stated goal is to make commanders "smarter" by organizing their information, enhancing wargaming capabilities, and giving them the tools to effectively synchronize operations. The assessment techniques available to judge the impact of these technologies tend to emphasize measurable outcomes – whether the technologies produce better kill ratios, quicker decisions, or the ability to analyze more information. However, outcomes are only part of the story. They do not tell us whether commanders are making "smarter" decisions or whether the technologies support the continued development of thinking skills and expertise development. We must also investigate the impact of technology on the higher-order thinking skills they purport to facilitate in order to improve performance.

The purpose of this effort was to create an assessment tool to measure the tactical thinking skills of officers in combat arms branches of the military. The product is a set of T-BARS intended to enable measurement of cognitive proficiency on tactical exercises by coding observable behaviors. While we have noted the applicability of such an assessment tool for the evaluation of advanced technologies, the focus of this effort was on evaluation in the context of training applications.

The remainder of this report is organized into four sections. In the first section, Perspectives Guiding the Development of the Assessment Tool, we describe the underlying perspectives and past efforts that formed the foundations of the current assessment tool development effort. In Development of the Tactical Thinking Behaviorally Anchored Rating Scales, we describe how the assessment tool was developed – the methodology for producing the final product. This section includes a discussion of how data was collected and analyzed to develop the scales, and how an analysis was conducted to measure the interrater reliability of the scales. In the final section, Discussion and Conclusions, we conclude with a discussion of how the T-BARS should be applied in practice to measure tactical thinking proficiency. This section provides an overview of the application of the tool, with reference to a separate User Guide that provides more comprehensive directions for usage. The assessment tool is contained in Appendix C.

Perspectives Guiding Development of the Assessment Tool

During their careers, officers amass an impressive command of declarative knowledge and procedural information, but this does not automatically lead to knowing how to make decisions and understanding situations during performance. To improve their cognitive skills and prepare for combat situations, officers should engage in deliberate practice in context-rich environments, including training scenarios, simulations, and field exercises. Deliberate practice is training that is structured to provide an opportunity to develop specifically targeted skills by practicing them and receiving feedback on performance. Previous research has shown that tactical thinking skills can be deliberately practiced and improved (e.g., Lussier, Ross, & Mayes, 2000; Lussier, Shadrick, & Prevou, 2003; Ross & Lussier, 1999; Ross, Phillips, Klein, & Cohn, 2005).

Techniques for evaluating the impact of training interventions involving deliberate practice typically require highly customized assessment tools and measures in order to quantify learners' improvements (e.g., Baxter, Harris-Thompson, & Phillips, 2004). It is costly to develop tailored measures for every new intervention, especially when these interventions are often scenario-based and require a distinct set of measures for each unique scenario. Furthermore, when customized evaluation measures are employed, it is challenging to compare outcomes across training interventions. A standardized assessment tool for tactical thinking skills that can be broadly applied enables us to compare and contrast relative values of tactical thinking trainers and support technologies while minimizing the cost of the evaluation.

Assessment of an individual's cognitive skills can serve purposes beyond gauging the effectiveness of a particular training implementation. It can also reveal a student's current aptitude in order to tailor training most effectively for that person. Training professionals have

little to no guidance for assessing or diagnosing aspects of the trainee's cognitive proficiency as part of implementation. It has been found that learners in complex cognitive domains such as tactical thinking respond best to interventions that incorporate instructional strategies and domain context appropriate to the learner's current level of cognitive proficiency (Ross, et al., 2005). Instructors and other training professionals could optimize their delivery of training with the use of a reliable diagnostic tool.

In light of these requirements, we set out to meet the following goals with the T-BARS assessment tool:

- Provide a standardized tool for assessing tactical thinking proficiency.
- Develop an assessment tool that can be used diagnostically for individual learners, and as a means to assess the impact of training interventions on tactical thinking skills.
- Develop a tool that can be applied to measure the impact of advanced technologies on user cognition (this was a secondary goal).
- Develop a tool that is not dependent on expert judgments, self-report, or intense interviewing and analysis to rate levels of tactical thinking proficiency.
- Support a user audience that is not highly specialized or experienced in assessing tactical thinking. The target audience for the T-BARS tool is researchers or professionals who are familiar with the role of complex cognition in tactical tasks, and who understand the combat arms domain.

Macrocognition and Mental Models

What is meant by "complex cognition" and "tactical thinking skills?" The Macrocognition framework provides a useful structure for understanding the types of higher-order thinking skills we are targeting with the assessment tool (Klein, et al., 2003). Macrocognition is a level of description of the cognition that occurs in naturalistic or field decision making settings. It is a complement to microcognition, which encompasses the elementary building blocks of cognition and is the primary focus of most laboratory researchers. Macrocognition consists of a set of critical cognitive functions and the processes that support those functions (see Figure 1). Skills such as sense making, problem detection, and attention management are critical to successful performance in high-pressure, high-stakes situations, and particularly in the situations that call for tactical thinking on the part of commanders. However, macrocognitive activities in themselves are not necessarily amenable to measurement. Because they are internal processes, they are invisible to the observer. If assessment relies solely on the outcome of the macrocognitive activities, the story is incomplete. Outcomes do not always accurately reflect the performance of the individual. Furthermore, the more interesting and useful component of performance for intervening and adjusting that performance is the thought process, interpretation, or rationalization that drives the outcome. In some cases the thought process is flawed but the outcome is acceptable. In other cases the thought process is sound but the implementation of the decision is suboptimal.

A core assertion of this assessment tool development effort is the idea that macrocognitive activities are enabled by an individual's domain mental models. Mental models

have a central role across the literature in cognitive psychology, expertise, instructional research, artificial intelligence, and systems control research. At the same time, the literature reflects a lack of agreement on the definition of mental model. Related terms such as schema, knowledge structure, conceptual model, and others cloud the issue further. Rouse and Morris (1986) in their review of the mental model literature, observe that the difference in the scope of definitions across disciplines most likely reflects inherent differences between open-ended tasks and well-defined tasks. For convenience, we provide a definition to orient the reader to this discussion: "A mental model is a representation of some domain or situation that supports understanding, reasoning, and prediction" (Gentner, 2002, p. 9683). Mental models also support action. These cognitive functions – understanding, reasoning, prediction, and action – are akin to the macrocognitive functions. Glaser and Baxter (2000, p. 2), state that "as learning occurs, increasingly well-structured and qualitatively different organizations of knowledge develop." They believe the development of competence is based on the acquisition of knowledge in a highly connected and articulated way through interactions with the environment, especially first-hand experiences. Each experience, and the knowledge that stems from it, is organized in the form of mental models.

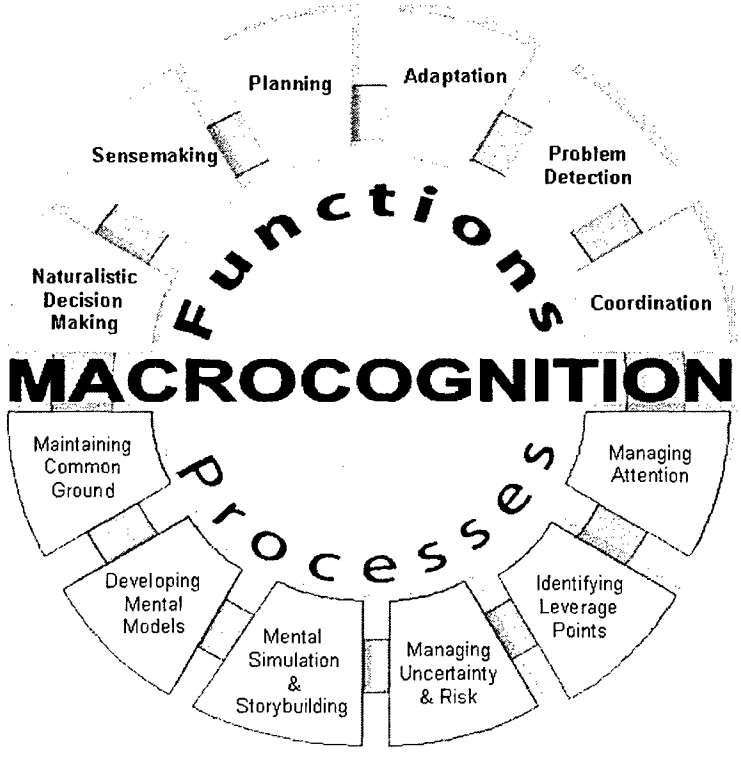


Figure 1. Macrocognition.

Trainers and instructors seek to improve macrocognitive abilities, whether or not they state it explicitly, through experiential training in the form of scenarios, vignettes, simulations, live field exercises, and the like. The goal is essentially to build an experience base in each individual that enables him or her to produce new knowledge about how complex cognitive tasks are accomplished in a specific domain or environment. The outcome of good experiential training is the broadening or deepening of mental models that enable macrocognition (and at times the replacement of faulty mental models with more accurate ones). If mental models

organize the knowledge and experience that is required to execute macrocognitive activities, then by measuring the depth and breadth of an individual's mental models we have a window into his or her thought processes and cognitive skills.

Think Like A Commander

In order to assess macrocognitive skills in the specific context of tactical thinking, we must succinctly define the domain-specific mental models. The Think Like A Commander (TLAC) research program (Lussier, 1998; Lussier et al., 2003) defines eight "themes" that expert commanders are thought to use on the battlefield. The themes were derived from interviews with numerous tactical experts (Deckert, Entin, Entin, MacMillan, & Serfaty, 1996) and represent mental models of tactical thinking or the cognitive processes experts use. A TLAC program of training was subsequently developed with the goal of training Soldiers and leaders to be better adaptive leaders by becoming proficient at the eight themes during deliberate practice. The TLAC training is currently in use at Fort Knox in the Armor Captain's Career Course, and in the Reserve Component Armor Captain's Career Course as a distance learning application.

The eight TLAC themes were utilized as the basis for the domain mental models to be measured with T-BARS. They are the following:

Know and Use All Assets Available (Assets). This theme refers to the necessity of combat leaders to maintain awareness of the synergistic effects of fighting their command as a combined arms team. This includes not only all assets under their command, but also those which higher headquarters might bring to bear to assist them.

Focus on the Mission and Higher's Intent (Mission). This theme refers to the need for leaders to always stay aware of the higher purpose and results they are directed to achieve. Even when unusual and critical events may draw them in a different direction, it is essential to stay focused on the overall mission.

Model a Thinking Enemy (Enemy). The focus of this theme is on the importance of remembering that the adversary is a reasoning human being who is intent on defeating friendly forces. Although it's tempting to simplify the battlefield by treating the enemy as static or simply reactive, this will harm the Soldier's ability to fight an effective battle.

Consider Effects of Terrain (Terrain). This theme reflects the importance of not losing sight of the operational effects of the terrain on which they must fight. Every combination of terrain and weather has a significant effect on what can and should be done to accomplish the mission.

Consider Timing (Timing). The focus of this theme is on the importance of being cognizant of the time available to get things done. A good sense of how much time it takes to accomplish various battlefield tasks and the proper use of that sense is a vital combat multiplier.

See the Big Picture (Big Picture). This theme refers to the importance of maintaining awareness of what is happening in the environment and how it might affect operations—what courses of action can affect others' operations. A narrow focus on one's own fight can result in a leader being blind-sided.

Consider Contingencies and Remain Flexible (Contingencies). Commanders must never lose sight of the old maxim that "no plan survives first contact with the enemy." Flexible plans and well thought out contingencies result in rapid, effective responses under fire. Contingencies are characterized by thinking that begins with questions like "What if...?" or "How else can I...?"

Visualize the Battlefield (Visualize). Leaders must be able to visualize a fluid and dynamic battlefield with some accuracy and use this visualization to their advantage. A leader who develops this difficult skill can reason proactively like no other.

Lussier and his colleagues generated general descriptions of the nature of performance along each of the eight TLAC themes as skill improves (Lussier, 1998). For example, related to the Mission, inexperienced tacticians tend to focus narrowly on their own mission. Highly experienced individuals, on the other hand, consider the objectives of the larger unit and are able to conduct their mission in a manner that supports the higher intent. Lussier's general descriptors, represented in Figure 2, provided the initial basis for the assessment tool developed in this effort.

Behaviorally Anchored Rating Scale (BARS)

Prior to this effort, the eight TLAC themes had been incorporated into an experimental assessment tool, to determine whether individuals' tactical mental models could be measured based on their observed performance in a tactical exercise. This experimental tool was developed as a BARS. Traditionally, BARS have been used in organizational settings to measure the effectiveness of individuals performing a wide variety of tasks (Muchinsky, 2003). A typical BARS lists observable behaviors that correspond to a numeric score, with higher numbers indicating more advanced behaviors. The BARS generally utilize five performance points with '1' representing a low level of performance and '5' representing a very high level of performance. To construct each scale, performance is observed in the work setting and/or incidents from these observations are gathered from subject-matter experts (SMEs). These incidents are placed along a scale with a range from poor to excellent. Once a BARS is developed for a particular task or job position, individuals without domain experience or expertise have a structure with which to rate performance by assigning scores to behaviors they observe. Figure 3 contains an example of a BARS for evaluating nurses.

The BARS format is appealing for assessing tactical thinking skills for two key reasons. First, it allows evaluation of invisible cognitive processes by categorizing them as overt behaviors. Second, it allows a means of judging proficiency without being an expert in the field. In previous research efforts where the experimental TLAC BARS have been applied, the BARS structure has shown great potential as a technique for measuring individuals' tactical thinking mental models (Phillips, Shafer, Ross, Baxter, & Harris, 2003; Ross, Battaglia, Hutton, & Crandall, 2003). However, this tactical thinking BARS, or T-BARS (see example shown in Table 1), required extensive modification and systematic testing to be utilized as a reliable assessment tool. Accordingly, the objective of this effort was to extend and refine and expand the scales for use by researchers and other experienced observer-controllers who wish to reliably measure tactical thinking mental models and performance on tactical decision tasks.

T1. Know and Use All Assets Available

Know Data About Systems	Link Systems to Mission Requirements	Dynamic Friendly Model	See Own Unit in Context of Larger Unit Assets	Command the Force

T2. Keep a Focus on the Mission and Higher's Intent

Focus on Own Missions	Discriminate Intent and Explicit Mission	Model Effect of Own Mission on Headquarters	Accurately Predict Impact of Own Actions	Support Intent

T3. Model a Thinking Enemy

Ignore Enemy	Use Enemy Templates	Model a Thinking Enemy	Accurately Predict Enemy Actions	Deny Enemy Intent

T4. Consider Effects of Terrain

OCOKA	Recognize Important Aspects	Dynamic Terrain Model	Use Terrain to Own Advantage	Shape the Battlefield

T5. Consider Timing

Ignore Timing	Aware of Timing Constraints	Model Time Against Assets, Terrain, Objective	Timely, Clear Orders	Bold Actions

T6. See the Big Picture

Focus on Own Actions	Understand Significance of Enemy/Friendly States/Events	Model Effect of States & Events On Battle	Act to Facilitate Larger Organization's Actions	Teamwork

T7. Consider Contingencies and Remain Flexible

Adhere to Plan	Consider Other Approaches	Recognize Leverage Points/Options	Model Different Approaches Against Enemy Model	Adaptability

T8. Visualize the Battlefield

Prioritize Events and Things to Attend to	Visualize Timing, Assets & Terrain	Visualize Potential Outcomes Accurately	Visualize Second & Third Order Consequences	Shape the Future Fight

Figure 2. General descriptors of the progression of tactical thinking across the Think Like A Commander Themes.

Note: OCOKA refers to Observation, Cover and Concealment, Obstacles, Key Terrain, Avenues of Approach.

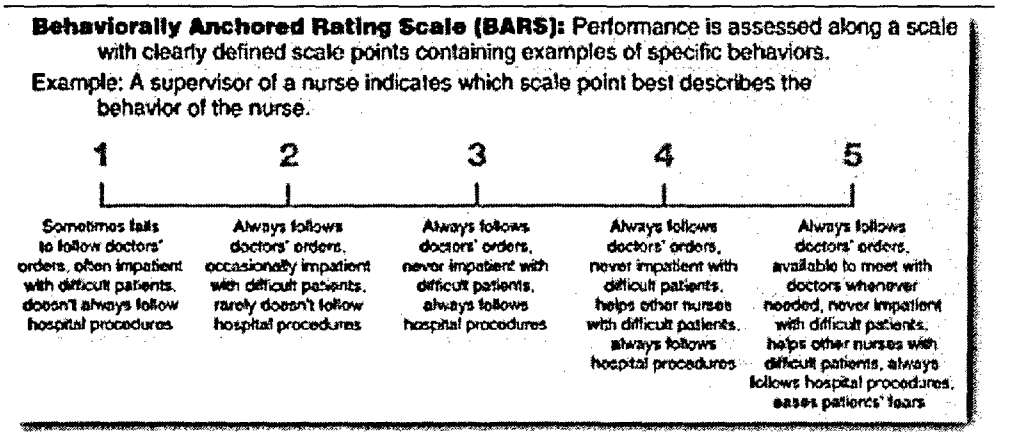


Figure 3. A sample BARS for nurses. From http://www.navarrocollege.edu/votech_programs/business/courses/bmgt1303powerpointforweb/bmgt1303chapter11powerpoint.htm#slide083.htm.

Table 1

Initial Experimental Version of the T-BARS for the Theme Know and Use All Assets Available

Focus on Own Mission	Discriminate Intent and Explicit Mission	Model Effect of Own Mission on HQ Intent	Accurate Predictions	Support Intent
1	2	3	4	5
•Articulates an understanding of the mission without any consideration of higher intent. •Neglects to keep HQ informed of plans and situation. •Neglects to request reinforcements when the plan requires it. •Ignores or loses sight of higher intent when distracted by unusual events.	•Can differentiate mission from higher intent, but does not apply these differences to understanding the current situation in front of him. •Understands both mission and intent, but does not consider whether mission will support that intent, or whether it needs to be modified in any way to better support intent.	•Considers whether mission will support the intent. •Considers whether mission needs to be modified in order to better support the intent. •Considers ways to modify mission to better support intent. •Thinks through what has to be accomplished in order for the higher intent to be achieved.	•Mentally simulates how his mission will contribute to achieving large mission. •Prioritizes what needs to happen in order for the higher mission to be accomplished (e.g., "I need to do this instead of that").	•Articulates how and/or why his plan or COA supports the commander's intent. •Allows intent and current situation to guide the COA rather than the explicit mission.

Note: COA refers to course of action.

Development of the Tactical Thinking Behaviorally Anchored Rating Scales

With the goal to refine and extend the existing T-BARS, researchers familiar with the tactical thinking domain focused on Army combat arms officers and their macrocognitive activities in the context of a range of tactical exercises. Data were collected through interviews with officers of varying ranks and experience levels. The range of performance exhibited in the data was examined to develop new behavioral descriptors within the T-BARS or refine existing descriptors. Updated versions of the T-BARS were tested against portions of the data set for

interrater reliability. This process continued iteratively until the T-BARS contained adequate descriptors for the entire range of performance (from level 1 to level 5) and proved reliable when applied by researchers not involved in its development. In this section we describe the process by which the T-BARS were extended, refined, and tested.

Interview Methods

Multiple data collection protocols were employed to elicit data from the Army officers who participated in the study regarding their thoughts and decisions related to tactical problems. The protocols were adapted over time as their effectiveness for eliciting the desired data became apparent. The key elements of each protocol, however, were the TLAC vignettes and the interviews.

Vignettes. A pool of six vignettes provided the tactical challenge to which participants responded, with most data collection sessions employing a subset of three vignettes. Each vignette was obtained from the TLAC program for training Army Captains, and placed the participant in the position of a company commander during a combat arms mission set in Azerbaijan. Participants read a Road to War background description and Operations Order or Fragmentary Order containing information specific to the mission. They were provided with a Rules of Engagement document and maps upon request. The vignettes themselves were Flash-based scenarios containing maps and graphics to indicate movements and locations. The vignettes were pre-scripted and evolved over time, with narration accompanied by incoming situation reports and other communications from characters within the mission (e.g., platoon leaders, local civilians, etc). The vignettes addressed a variety of operational challenges. They were:

> *Vignette 1: Establish a Safe Route.* The participant is required to clear a route through potentially hostile country into an urban area, accompanied by an assistant to a US ambassador. The participant must decide on a route to the objective and determine how to handle his interaction with the ambassador's assistant, whose objectives are not aligned with the company's mission.
>
> *Vignette 2: Enable Humanitarian Operations.* While escorting a humanitarian aid convoy to a refugee camp, the participant comes upon a flooded village in need of help. The participant has to weigh his ability to complete the original mission against the pop-up opportunity to help the villagers. He must also predict the impact his actions will have at the site of the flooded village.
>
> *Vignette 3: Man a Border Outpost.* The participant controls five border outposts. In the midst of a holiday celebration, an explosion occurs and one outpost no longer responds to communications. The participant must assess the source of the explosion and determine the appropriate level of force to employ in response.
>
> *Vignette 4: .Conduct Presence Patrols.* The participant's company is tasked with providing security in an area where civilians are returning to their homes. The participant is forced to determine what to do when a subordinate detains a group of men whose intent

is unclear. He must determine how to apply the rules of engagement and assess the intent of the detainees.

Vignette 5: Control a Civil Disturbance. The participant is required to handle a situation in which two opposing crowds form at a bridge that he is tasked to guard. A UN representative becomes involved in attempts to appease the crowds, and an embedded media crew is recording the incident. The participant must determine how to diffuse the situation.

Vignette 6: Destroy a Defeated Enemy. The brigade is pursuing a withdrawing enemy and the task force commander directs forces to halt and establish a hasty defense for the night. However, the participant's unit senses an immediate opportunity to attack and destroy a disorganized enemy unit.

Think-aloud protocol. The think-aloud protocol was derived from a technique developed by Klein, Phillips, Battaglia, Wiggins, and Ross (2002). Within each vignette, the participant was told, "Please think aloud about your responses to the following questions. What's important in this scenario? What information do you need? What will you do now?" If the participant fell silent at any point, the interviewer asked him or her to continue thinking aloud.

Simulation interview protocol. The Simulation Interview (SI) protocol was based on Militello and Hutton's (1998) Applied Cognitive Task Analysis technique. The SI developed by Militello and Hutton is intended to give the interviewer a better understanding of participants' cognitive processes in the context of an incident. In our case, the TLAC vignettes provided the incident. The SI consists of a number of probes about different aspects of the incident. The probes we used were tailored for each of the three stopping points in the vignettes. They focused on what participants perceived as important, why they noticed those things, how they saw the situation developing, their priorities, and what information they sought and why.

Group vs. individual interviews. Group interviews were conducted in the first data collection effort. During these sessions, groups of two to six participants were exposed to a vignette in its entirety. Each participant presented a response to the vignette, which was then discussed by the rest of the group. The interviewers then facilitated a group discussion of the vignette, focusing on how the participants interpreted the information provided by the vignette and utilized that interpretation to determine suitable actions.

Data Collection

Four primary rounds of data collection were conducted, one at Fort Campbell, one at Fort Carson, one at Fort Sill, and one at Fort Hood. Incidental interviews were also conducted at the School of Advanced Military Studies at Fort Leavenworth and at Fort Knox. After each round, the data were assessed and the protocol was refined. The intent was to collect data from Lieutenants, Captains, Majors, Lieutenant Colonels, and Colonels in order to populate all levels of the T-BARS, from novice to expert levels of performance. Participants' ranks should roughly correlate with levels of proficiency on tactical thinking tasks. While the correlation was not

calculated, this participant pool was anticipated to provide a balanced view of the range of performance. See Table 2 for a summary of participant ranks.

Table 2

Participant Ranks

	Fort Campbell	Fort Carson	Fort Sill	Fort Hood	Other	Total
Lieutenant	0	1	0	0	0	1
Captains	4	4	8	8	0	24
Major	4	8	8	6	1	27
Lieutenant Colonel	4	5	8	5	0	22
Colonel	0	0	0	0	2	2
Total	12	18	24	19	3	76

Fort Campbell. Twelve participants were interviewed in the first round of data collection. Their specialties ranged from infantry to chaplain. Group interviews were conducted for two vignettes, and the think-aloud protocol was utilized with individual participants for a third vignette. The order of the vignettes (#1, #2, and #3) was counterbalanced across groups. Participants were also asked to make a list of tactical thinking skills in order to compare their lists to the TLAC dimensions. The initial intent was to use the group session data to generate new behavioral descriptors for the T-BARS dimensions, and then rate the individuals' think-aloud data to check interrater reliability.

The group interview technique proved to generate less information and insight into cognitive processes than the individual think-aloud interviews. Furthermore, the responses tended to be amalgams of the group's thinking rather than genuine, organic responses from a single individual's thought process. After examining the data, we decided to use only the individual interview data. Further, we discovered that the think-aloud data was not as rich as was necessary to develop the behavioral descriptors in the T-BARS. This outcome could have been due to either fatigue, since individual interviews occurred at the end of four-hour sessions, or to the protocol itself.

Fort Carson. Eighteen participants were interviewed in one-on-one sessions during the second round of data collection. Both think-aloud and SI protocols were employed. Two initial prompts were used, one action-oriented (e.g., What would you do in this situation?) and one not action-oriented (e.g., What do you need to consider in this situation?). Vignettes 1, 2, and 3 were counterbalanced across participants. The interviewers also counterbalanced for prompt and the two interview types. In addition, each vignette was paused at pre-selected pivotal points, and the protocol was implemented in order to generate an understanding of how the participants' thinking about the vignette changed as the situation developed.

The action- or non-action-oriented prompts made no discernible difference in participant responses. With regard to protocol effectiveness, the SI tended to elicit richer information than the think-aloud technique. When asked to think aloud, participants tended to describe an action

plan and a few important items of information, but did not expand on their thought processes or the reasons behind their plan.

Fort Sill. Twenty-four participants were interviewed in the third round of data collection. Participants were interviewed individually using the SI protocol. Vignettes 1, 2, and 3 were counterbalanced across participants.

Initial analysis of the interview data indicated that vignettes 1, 2, and 3 were not producing the distribution of data needed to fill out all of the T-BARS themes. For example, the terrain in the scenarios was not represented at a high degree of granularity, and the situation was not designed to encourage participants to thoroughly assess the impact of terrain on their mission (although some of the higher performing participants did exhibit significant consideration of the terrain). As a result, the data were not revealing a suitable quantity of behavioral descriptors within the *Terrain* theme. Therefore, we decided to use three new vignettes in the next round of data collection – #4 and #5.

Fort Hood. In the fourth round of data collection 19 participants were interviewed. On the first day of interviewing, participants responded to vignettes 4, 5, and 6. However, vignette 6 did not yield as much tactical thinking data as the others, so interviewers conducted vignettes 1, 4, and 5 (counterbalanced) on the second and third days of data collection.

Other. Interviews were conducted at Fort Knox with a recently retired colonel, and at the School of Advanced Military Studies at Fort Leavenworth with a colonel and a major. In addition to the data collected with these participants, archival data from lieutenants (to represent the early stages of tactical thinking skill) and generals (to represent mature tactical thinking skills) involved in exercises or incidents were utilized in the sample to fill out the full range of behavioral descriptors within the T-BARS. Note that the archival data were not generated using TLAC vignettes, as was true for most of the data applied to develop the T-BARS assessment tool.

Theoretical Underpinnings of the T-BARS

Analysis began with an inspection of the experimental version of T-BARS for internal consistency. Two researchers examined each descriptor in each theme and compared them to the other descriptors in that theme. The rating descriptions were revised within each theme and rearranged to create a more uniform and consistent progression within the theme. The intent was to reduce the potential for confusion on the part of the T-BARS user and prevent multiple interpretations as much as possible.

After working with the original T-BARS and the data collected, it became clear that the T-BARS required a solid theoretical grounding for its five-step progression. The cognitive psychology, expertise, training, and education literatures were examined for candidate frameworks to guide the characterization of performance and behavior at different levels of the T-BARS. For example, Bloom's taxonomy (Bloom, 1956) was considered for its descriptors of how individuals develop and apply their knowledge as they become more proficient in a domain. However, the Dreyfus and Dreyfus (1986) five-stage model of skill acquisition was deemed a

more appropriate framework for the T-BARS tool, as it specifically pertains to domains like tactical thinking that are ill-structured and cognitively complex.

The Dreyfus and Dreyfus (1986) five-stage model of skill acquisition characterizes five performance levels through which individuals progress as they gain skill and proficiency in cognitively complex domains: novice, advanced beginner, competent, proficient, and expert. The model has been applied to training and instruction within domains such as combat aviation, nursing, industrial accounting, psychotherapy, and chess (Benner, 1984; 2004; Dreyfus & Dreyfus, 1986; Houldsworth, O'Brien, Butler, & Edwards, 1997; McElroy, Greiner, & de Chesnay, 1991). Like tactical thinking, these domains demand that decisions be made quickly in environments that are complex, ambiguous, and dynamic. Further, skill can be acquired only through first-hand experience doing the task. The Dreyfus and Dreyfus (1986) model provides an excellent general structure that can be applied to describe levels of tactical thinking proficiency. The following is a summary of each of the five stages delineated in the model.

Stage 1: Novice. Novices have limited or no experience in situations characteristic of their domain. They exhibit rigid adherence to rules they have been taught, or plans they have been given. They have little situational perception, and they lack the basic domain knowledge needed to perform analysis.

Stage 2: Advanced Beginner. Advanced beginners have enough domain experience that their performance is marginally acceptable. They have a sufficient knowledge base with which to analyze a situation. At this stage they are able to recognize recurring, meaningful "aspects" of situations—global characteristics identifiable only through prior experience where the prior experience provides a comparison case for the current situation. Their knowledge base regarding aspects and attributes of situations enables them to develop their own guidelines for action. However, all components of the situation tend to be treated as independent pieces and as equal in importance, rather than differentially weighted based on the circumstances and goals.

Stage 3: Competent. At the competent level, performers have mental models that they can apply to new situations. This stage is marked by the ability to envision and predict how a situation is likely to play out, which guides the formulation, prioritization, and management of longer-term goals. Competent performers are very planful, where advanced beginners are more reactive. However, competent individuals tend to adhere to the plan as the situation plays out, even when circumstances change. They have difficulty adapting their plan to address new situational demands.

Stage 4: Proficient. Proficient individuals' performance shifts from being guided by the plan to being responsive to the situation. They see the situation as an inseparable whole rather than as independent attributes; they have the ability to recognize meaningful patterns of cues without breaking them down into their component parts for analysis. As such, they are able to intuitively assess what is happening and what is most critical for achieving success. They shift their assessment of the situation as it evolves and changes, and they can adjust their course of action accordingly. However, while their situation

assessment is recognitional and intuitive, they still perform deliberate analysis when making decisions and devising or adjusting a course of action.

Stage 5: Expert. Expert performance is marked by a shift to recognitional decision making. Experts intuitively assess the situation and also intuitively recognize a suitable course of action that will accomplish their goals. They have a substantial base of experience from which to operate. Their mental models are broad, deep, and elaborate. They are able to make fine discriminations between perceptual cues (Klein & Hoffman, 1993), and can diagnose and assess situations that confuse or stump their less-experienced peers. Experts also have a wide range of routines and tactics for getting things done (Klein, 1998).

The five stages of the Dreyfus and Dreyfus (1986) model readily mapped onto the 5 levels seen in the general descriptors of tactical thinking performance (shown previously in Figure 2) provided by Lussier (1998) for each of the TLAC themes. Lussier had articulated a progression of tactical thinking skills specifically as observed in his research and training of tacticians. The Dreyfus and Dreyfus stages describe the progression of cognitive skill development in general, independent of domain. The value of applying the five-stage model to the T-BARS is that it provides a cognitive profile that can anchor the development and refinement of the domain-specific descriptors in the T-BARS. Table 3 provides an example of the Stage 3 cognitive profile, incorporating characteristics of knowledge and performance exhibited by competent performers. The full listing of knowledge and performance characteristics for each of the five stages is provided in Appendix A. As our tactical thinking data were parsed and developed into behavioral descriptors, the descriptors were assessed against the Dreyfus and Dreyfus cognitive profiles as a means of ensuring that they were placed at the appropriate level (category 1, 2, 3, 4, or 5) in the scales.

We hypothesized that the themes representing mental models – Assets, Mission, Enemy, and Terrain – must be built up to some basic level of comprehension before the themes representing cognitive processes – Timing, Big Picture, Contingencies, and Visualization – can be implemented (Ross et al., 2003; Ross, Battaglia, Phillips, Domeshek, & Lussier, 2003).

Figure 4 illustrates this hypothesized developmental process. The themes representing cognitive processes are exhibited by experienced, proficient tactical decision makers. They conduct these higher-order mental operations in the context of the basic mental models represented by the first four themes. For example, an experienced tactician can estimate how long it will take to move a bridging asset from one point to another (*Timing* in the context of *Assets*) or predict what the enemy will attempt as the situation plays out (*Visualization* in the context of *Enemy*). Accordingly, the T-BARS tool was refined by incorporating the behaviors associated with the cognitive process themes into the mental models themes, thereby resulting in four T-BARS (*Assets*, *Mission*, *Enemy*, and *Terrain*) rather than eight.

Table 3

Cognitive Profile for Stage 3: Competent Individuals

STAGE 3: COMPETENT	
General Characteristics	
Knowledge	***Performance***
•How to think about the situation in terms of overarching goals or tasks (Benner, 1984). •The relative importance of subtasks depending on situational demands (Benner, 1984; Dreyfus & Dreyfus, 1986). •Particular patterns of cues suggest particular conclusions, decisions, or expectations (Dreyfus & Dreyfus, 1986). •A personalized set of guiding principles based on experience (Houldsworth et al., 1997). •How to anticipate future problems (Houldsworth et al., 1997).	•Is analytic, conscious, and deliberate (Benner, 1984; Dreyfus & Dreyfus, 1986). •Does not rely on a set of rules (Houldsworth et al., 1997). •Is efficient and organized (Benner, 1984; Dreyfus & Dreyfus, 1986). •Is driven by an organizing plan that is generated at the outset of the situation (Dreyfus & Dreyfus, 1986). •Reflects an inability to digress from the plan, even when faced with new, conflicting information (Dreyfus & Dreyfus, 1986). •Reflects an inability to see newly relevant cues due to the organizing plan or structure that directs attention (Benner, 2004). •Reflects an emotionally involved performer who takes ownership of successes and failures (Dreyfus & Dreyfus, 1986). •Focuses on independent features of the situation rather than a synthesis of the whole (Houldsworth et al., 1997).

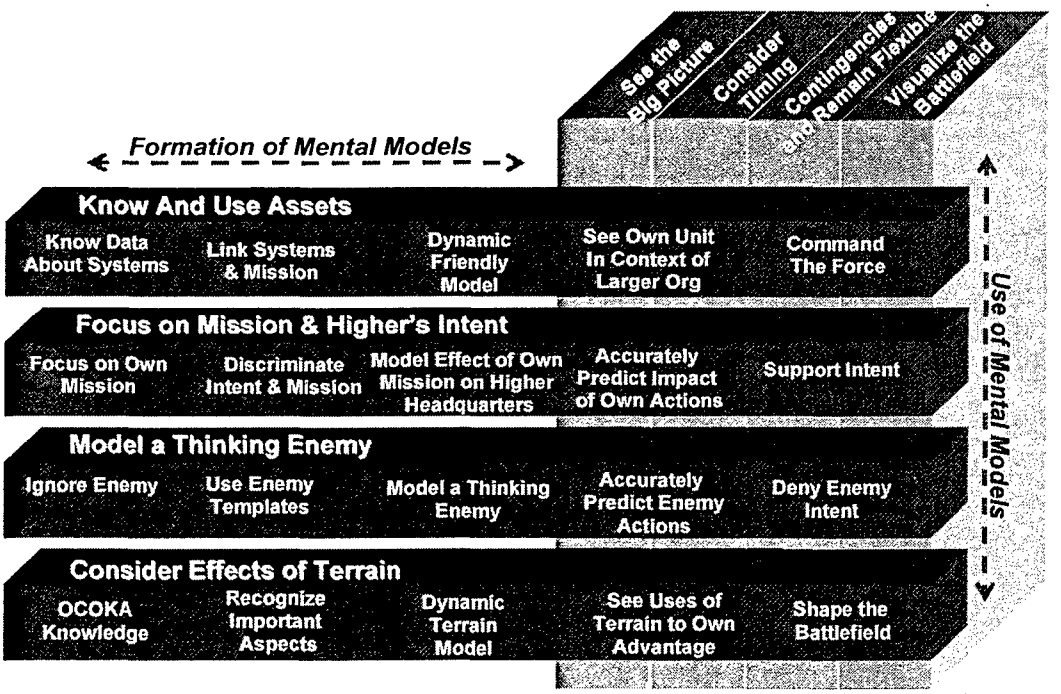


Figure 4. Hypothesized developmental sequence of tactical mental models and cognitive processes.

Revisions of the T-BARS

The T-BARS underwent several revisions over the course of the effort, iterated with interrater reliability testing. The researchers began with the experimental version (see for example Table 2) of the T-BARS as a starting point from which to add, delete, and modify behavioral descriptors within the scales. A subset of the data from each round of interviews was examined. In some cases, behavioral descriptors in the existing T-BARS matched the behaviors exhibited in the data. In other cases, new behavioral descriptors were generated to account for the participants' behaviors and thought processes. As would be expected, more behavioral descriptors were newly generated at the beginning of the revision process than toward the end.

In generating the new behavioral descriptors, the researchers attempted to generalize the descriptors to the extent that they could be used to describe a range of similar behaviors that may be found in other data records. For example, one participant responded to a question about how to reach the objective by saying, "...the ground is pretty soft right now, so I would kind of reject [Route] Orange out of hand because it goes through the middle of a marsh." The behavioral descriptor generated for this data chunk was, "Rejects a route due to terrain conditions."

New and existing behavioral descriptors were placed into the T-BARS themes (e.g., *Assets*, *Mission*, *Terrain*, *Enemy*) based on which of these aspects of the tactical picture they most closely addressed. This judgment was straightforward. The judgment about the level in which to place a behavioral descriptor was guided by the cognitive profiles provided by the Dreyfus and Dreyfus five-stage model. As part of this process, tactical thinking profiles were

generated for each level of each T-BARS theme. In other words, the general cognitive profiles derived from the five-stage model were adapted into domain-specific profiles. Table 4 provides an example of the tactical thinking profiles within the Assets theme. The profiles for all themes are available in Appendix B.

Table 4

Tactical Thinking Profiles within the Assets Theme

Know and Use All Assets Available. Combat leaders must not lose sight of the synergistic effects of fighting their command as a combined arms team – this includes not only all assets under their command, but also those which higher headquarters might bring to bear to assist them.				
Knows Textbook Capabilities	Matches Assets to Mission Requirements	Utilizes Organic Assets to Accomplish Mission Objectives	Recognizes Full Range of Assets Required based on Situational Demands	Applies Full Range of Assets to Direct the Outcome of the Battle
1	2	3	4	5
Performance is abstract and rule-based, and focuses on variables in isolation. Individual knows facts about standard capabilities of organic assets such as ranges of weapons, number of vehicles per unit, and so forth. The foundational knowledge required to analyze how assets can be applied to the situation has not yet developed.	Performance reflects simple analytical processing using a limited experience base. Organic assets are matched to mission requirements. For example, a tank formation would be allocated to the area where heavy armor is needed for protection. Individual has difficulty prioritizing tasks, so asset utilization is driven by capabilities (what the asset can do) over situational demand (what is the most pressing mission task).	Performance reflects a mental model of asset utilization, but remains dependent on analysis and planning rather than recognition and intuition. Individual can prioritize mission tasks and predict how the situation could unfold, and an asset utilization plan is generated against that analysis. However, execution is driven by the plan over the situation, so individual has difficulty adjusting asset utilization to meet changing situational demands.	Performance reflects a recognitional or intuitive assessment of the situation, but analytical decision making where the individual deliberates about a course of action. Individual recognizes the availability of non-organic and non-military assets in addition to his own organic assets. For example, civilians are recognized to be valuable sources of human intelligence (HUMINT). Situational demands drive asset utilization, rather than the plan or the organic assets at the individual's disposal.	Performance reflects a recognitional ability to assess and decide. Individual can visualize specific outcomes of asset utilization and has the ability to avoid unwanted consequences. For example, he knows how to command and maneuver his forces to avoid an uprising by the locals. Individual leverages and coordinates organic, non-organic, and non-military assets to achieve mission objectives.

Final Review of the T-BARS

Once profiles had been generated for each level of tactical thinking within each of the four themes, and behavioral descriptors had been defined to account for the interview data samples, the next step was to conduct a final review of the T-BARS prior to conducting a final round of interrater reliability testing. The task was to ensure a consistent pattern of behavioral descriptors within each theme. Specifically, the review:

- Identified "absence of behavior" descriptors and reworded them into observable performance statements.
- Ensured that performance statements in one level were addressed in the next level so that an improvement in performance was reflected as the levels progressed.
- Ensured that descriptors were specific, observable behaviors rather than general statements.
- Reworded descriptors so that every one began with a verb to indicate observed behavior. (The http://www.officeport.com/edu/blooms.htm website was consulted as a job aid to suggest verbs appropriate for different levels of cognitive performance in Bloom's Taxonomy.)
- Revised items for increased clarity and simplicity.
- Examined the scales for indications in the descriptors that another mental model or a cognitive process (e.g., *Big Picture*, *Timing*, *Contingencies*, and *Visualization*) was being considered as the primary behavior performed, and document the links between the scales accordingly.

The review process allowed a check for integrity (face validity) of descriptors with regard to: (1) the tactical thinking profile, which is a description of the general performance of that rating level for that theme; (2) other descriptors within the scale, to ensure their consistency and avoid conflicts amongst them; (3) other descriptors of that rating level for the other themes; and (4) the general description of that rating level (according to Dreyfus & Dreyfus [1986]). Key trends we looked for were (1) as levels of performance progressed, *Big Picture*, *Timing*, *Contingencies*, and *Visualization* were indicated more often and more often in combination, and (2) as levels of performance progressed, mental models were found to more often work in concert. The four finalized T-BARS can be found in Appendix C.

T-BARS User Guide

Following the development of the T-BARS and the interrater reliability testing, a User Guide was generated for researchers who will implement the assessment tool (see Phillips, Ross, & Shadrick, in preparation). The user guide consists of the following:

- Background information about tactical thinking mental models and how they can be measured using the T-BARS tool.
- Tactical thinking profiles for each level within each theme.
- Instructions for implementing the T-BARS as an assessment tool, including how to rate performance and how to score ratings.
- Instructions for interpreting the scores generated from the T-BARS tool.

- Guidance for achieving interrater reliability within a group of researchers utilizing T-BARS.

Interrater Reliability

Three rounds of interrater reliability testing were conducted during the T-BARS development effort. After each round, the T-BARS underwent further scale development and refinement based on the results of the ratings. We discuss each round in turn.

Round One. In the first round of reliability testing, three individual vignette responses were selected as the training sample. These vignettes represented weak, average, and strong responses in order to be able to rate the widest possible range of responses. We began by rating each participant response as a whole unit, giving one score for each of the eight themes for each participant. Rating at the level of a vignette response turned out to be too much of a leap, requiring too much domain knowledge on the part of the researcher and giving too much latitude for inference and multiple interpretations. A second pass through the data was then made by rating data segments, or small sections of the text, as well as individual sentences. Rating sentences proved difficult as in many cases a thought is only partially expressed in a single sentence. Ratings of data segments, however, was effective. Each segment reflected a single thought or consideration by the participant, and as such was amenable to rating using the T-BARS.

Five researchers rated the three vignettes which segmented into 52 items. Percent agreement for theme was compared for combinations of raters. At this stage in development, all eight TLAC themes were represented in the T-BARS tool. Agreement among the five raters for theme was at 15.4%. At least three raters agreed on the applicable theme for an item 75% of the time. Pairs of raters were sampled and it was found that agreement at the theme level varied between 40% and 50%.

After rating the themes, two of the five researchers conferred in order to reach agreement on theme for every data segment. The two researchers then independently rated the level (1-5) of each data segment. Agreement between the two raters on level once theme agreement was reached was 57.7%.

Round Two. The second round of testing was conducted with a focus on the question of whether the behavioral descriptors in the T-BARS successfully captured the entire range of behaviors represented in the data set, across vignettes and across experience levels. For that reason, statistics were not calculated.

Two researchers coded two participant transcripts using an improved version of the T-BARS. The transcripts were chunked into 103 data segments. For each segment the rater assigned a theme and level, and annotated his or her ratings with notes describing the rationale for the ratings. In cases where the existing behavioral descriptor did not adequately capture the behavior represented in the data segment, raters either re-worded the descriptor to broaden or clarify it, or generated a new descriptor to account for the data. The researchers then compared ratings, adjusted behavioral descriptors, and generated new descriptors.

The behavioral descriptors recommended by Round Two raters were synthesized and incorporated into the scales. The T-BARS product resulting from Round Two adjustments was later subjected to a review (described above as the Final Review of the T-BARS) by a third researcher for internal consistency of the themes or mental models of tactical thinking, and the levels of performance within each theme. That is, the reviewer compared all the descriptors within each level, 1-5, across all the themes, to ensure their consistency with regard to cognitive proficiency and their reflection of the stages of performance set forth by the Dreyfus and Dreyfus (1986) stage model.

Round Three. In the third round of testing, raters who had not been involved in the T-BARS development effort coded interview data. They employed the finalized version of the T-BARS, which consisted of the four themes deemed to represent tactical mental models (Assets, Mission, Enemy, and Terrain). We have previously articulated that the target audience for the T-BARS assessment tool is military researchers with at least a moderate education or experience base in the field of applied psychology. Researchers using the T-BARS tool should not be required to be specialists in the cognitive aspects of tactical decision making in order to apply T-BARS effectively. Accordingly, for the final round of reliability testing we sought raters with a moderate degree of experience conducting applied cognitive research, and moderate familiarity with the combat arms domain. Three raters were selected, each of whom had at least three but not more than five years of relevant experience.

An initial round of ratings was conducted in order to calibrate the raters to the technique and familiarize them with the scales for each theme. In this calibration round, 21 data segments were rated. The data segments were taken from an interview conducted during the Fort Sill data collection. This interview was deemed by the researchers to contain good variation on the themes represented, and reasonable variation on the levels. Variability was desirable in the calibration round so that raters would have the opportunity to apply a wide range of behavioral descriptors. For each data segment, the raters independently indicated the theme, the level within that theme, and the behavioral descriptor within that level that accounted for the content of the data segment. The complete protocol for the calibration coding is documented in Appendix D.

Once the calibration coding was complete, the raters met with a researcher who had developed the T-BARS to review the ratings and discuss problems or uncertainties. The protocol was judged by all three raters to be straightforward and easy to follow. Two minor process adjustments were made as a result of the calibration round experience. First, raters found that the context surrounding the data segment in some cases had an impact on their ratings. For example, the interviewer's comments or the interviewee's utterances immediately before or after the segment in question could have bearing on the rating. It was determined that while context is important to understanding an individual's mental models, for the purposes of measuring interrater reliability there is a need for each rater to judge items consistently. Therefore raters were instructed to judge each data segment as a distinct item, without considering any surrounding contextual cues. Second, raters found that certain data segments seemed to contain more than one thought, and therefore broke those segments into two or three items and assigned them separate ratings. However, raters varied in their determinations of which segments should be dissected. As a result, most dissected segments could not be compared across raters; one rater

would assign two or three values within the segment, and the other raters would assign only one value. Raters were therefore instructed to inform each other (and the coordinating researcher) when they wished to divide a segment into multiple items, and thereby raters coded identical segments throughout the data set.

The adjusted protocol was applied to a new data set which served as the test data. The test data comprised portions of three separate interviews representing three distinct vignettes and three levels of interviewee experience. The coordinating researcher purposefully selected interview data from one very experienced tactician, one tactician with an intermediate level of experience, and one relatively inexperienced individual. The coordinating researcher divided the transcripts into 58 data segments to be coded by the raters.

Just as in the calibration coding round, raters independently coded the data segments by assigning each a value for theme, level, and specific behavioral descriptor. After adjusting for dissected segments, 64 data segments were rated by two or more raters and subjected to interrater reliability testing.

Statistical analyses of the data tested for interrater reliability both in terms of scale *consistency* and rater *consensus* (e.g., Stemler, 2004), on theme as well as level ratings. A Cronbach's alpha coefficient was computed to assess the consistency of the ratings. This test is useful when more than two judges have scored the data. It measures the extent to which the judges' ratings hold together to measure a common dimension (Stemler, 2004). An alpha value greater than .70 is considered acceptable, where the majority of the variance in ratings is due to true score variance rather than error variance. The theme ratings in our sample produced a Cronbach's alpha value of .84 (N=56), and the level ratings produced an alpha value of .80 (N=55).

Next we computed percent agreement between pairs of raters, in order to assess their consensus. With regard to themes, we found Rater #1 and Rater #2 to be in strong agreement (80%), while the other two pairings showed only moderate agreement (63% for Raters #1 and #3, and 52% for Raters #2 and #3). The average percent agreement across the three pairs was 65%.

The theme ratings revealed a disproportionate use of the *Mission* theme by Rater #3, with 41% of the items scored as *Mission* versus 34% for Rater #1 and 25% for Rater #2. This finding is not surprising. The overarching mission objectives typically guide the thinking of tacticians throughout tactical exercises. The mission provides a goal set that influences one's consideration of how to utilize assets, leverage terrain, and view the enemy. As such, it is reasonable that a rater would consider the *Mission* theme to be broader in scope than intended by the developers of the T-BARS. We judged that Rater #3 was in fact exercising a broader definition of the *Mission* theme than Raters #1 and #2. This led us to revisit the content of the *Mission* scale and revise it to more clearly distinguish the boundaries with the other three themes.

To judge consensus on level ratings, the level values were examined separately for cases where pairs of raters agreed on theme and therefore were selecting a behavioral descriptor from identical option sets, and for all cases regardless of agreement on theme where in some cases

raters were judging level using dissimilar behavioral descriptors. Consensus would be expected to be higher when raters agreed on theme than when they did not. However, if the level descriptors are consistently differentiating stages of cognitive proficiency regardless of the specific theme or mental model, then consensus should be reasonable even when raters did not agree on theme. This is exactly what we found. Percentage agreement on level for each pair of raters is shown in Table 5. Agreement was calculated for exact consensus on level, where each rater selected the same value on the scale from 1 to 5, as well as for one-point differentials, where raters disagreed by one point on the 5-point scale. Following the theme agreement results, Raters #1 and #2 also had the highest pair-wise agreement on level. When they agreed on theme, 79% of the time they either agreed on level or differed in category by one point. When Raters #1 and #2 did not agree on theme, they agreed on level or differed by one point 78% of the time.

Table 5

Interrater Consensus on Theme and Level

Rater Pair	Agreement on Theme	Agreement on Level When Agreed on Theme			Agreement on Level Independent of Theme Agreement		
		Exact Level Agreement	1 Point Differential	≤1 Point Differential	Exact Level Agreement	1 Point Differential	≤1 Point Differential
Rater #1 & Rater #2	80% (N=59)	36% (N=47)	43% (N=47)	79% (N=47)	36% (N=59)	42% (N=59)	78% (N=59)
Rater #1 & Rater #3	63% (N=59)	41% (N=37)	41% (N=37)	81% (N=37)	31% (N=58)	40% (N=58)	71% (N=58)
Rater #2 & Rater #3	52% (N=56)	21% (N=29)	48% (N=29)	69% (N=29)	20% (N=56)	43% (N=56)	62% (N=56)
Averages	65%	32.7%	44%	76.3%	29%	41.6%	70.3%

Discussion and Conclusions

The product of this effort is a reliable assessment tool that provides insight into the mental models, and thus the macrocognitive skills, of tactical decision makers. We set out to develop a standardized tool that would enable assessment of complex cognition in the tactical thinking domain without reliance on expert judgment, in-depth interviews and analyses, or highly-specialized researchers. T-BARS users do not have to infer combat leader's thoughts to judge macrocognitive skills; they can simply observe actions and utterances. The T-BARS tool successfully categorized the behaviors exhibited by tactical decision makers across the range of performance to the ordinal level of measurement. When applied, it enables users to describe a learner's current level of cognitive proficiency with regard to four mental models that provide a basis for battlefield decisions and judgments.

While the T-BARS has progressed significantly as a usable tool from its original experimental version, it is prudent to describe its boundary conditions for use as well as steps

that could broaden the scope of its implementation in the future. Following is a discussion of the ideal qualifications of T-BARS users, the uses for which we believe T-BARS is suited, and directions for future development.

Users of the T-BARS

The target audience for the T-BARS assessment tool was stated as researchers or highly experienced observer-controllers who are familiar with naturalistic cognition and military contexts. The results of the interrater reliability testing support this contention. Users must have a basic understanding of how cognitive processes such as sensemaking and problem detection are exhibited in practice by tactical decision makers. The behavioral descriptors in the T-BARS define what the user may observe or hear from the tactical leader, but it is necessary to understand the language of the tactician in order to make the linkage to a behavioral descriptor from the scales. The tactical language consists of numerous acronyms that become a part of fluid speech. It contains unique terms such as "phase line," "avenue of approach," and "area of responsibility" that must be readily understood. Further, it incorporates specialized definitions of words with corresponding implications – for example, to *task organize* a unit as an attachment means that it falls under the command and control of the unit to which it is *attached.* A researcher utilizing T-BARS must be able to understand the associations tacticians are making within their specialized vocabulary in order to accurately judge what is being observed.

While the individuals most likely to have an appropriate background for use of the T-BARS tool are researchers, we have also seen that some instructors have an appreciation for the cognition that drives performance. These instructors may also be successful in using T-BARS to measure the performance of their students in tactical exercises.

It is our recommendation that T-BARS users work in pairs, especially during initial usage of the tool, to calibrate their application of the behavioral descriptors. While we have constructed the scales to be as precise and unambiguous in their descriptions of behaviors as possible, there remains some degree of variability in interpretation simply due to the nature of the instruments. Suggested techniques for calibrating across raters can be found in the *T-BARS User Guide* (Phillips, Ross, & Shadrick, in preparation).

Uses of the T-BARS

We envision two broad areas – training and technology evaluation – for which the T-BARS can provide valuable input regarding cognitive performance and application of mental models for a particular task. There may be other applications of the tool that we have not considered at this time. Below we discuss the ways in which the T-BARS could be implemented for these two instances.

Training. With regard to assessment in the context of training, T-BARS provides a means of measuring an individual's tactical thinking skills. The results of a T-BARS assessment can provide meaning in several ways. First, an individual's cognitive performance can be tracked over time to determine whether he or she is changing as a result of a training intervention or a real-world experience. Second, an individual's cognitive proficiency can be diagnosed in

order to determine the optimal course of instruction to develop him or her into a well-rounded tactical thinker. Finally, a training intervention can be evaluated on the basis of how individuals' cognitive performance is impacted over the course of the training.

It is possible that T-BARS could also measure team cognition on tactical tasks, although this was not the original intent and we have not attempted to employ the T-BARS in a team setting. The T-BARS tool might adequately capture a portion of a team's cognitive performance on a tactical thinking task, however critical aspects of the team mind such as common grounding and defining roles and functions would not be addressed by the assessment. It is likely, however, that a BARS-like scale could be developed to do just that – evaluate the quality of the team mind for a particular group of individuals working collaboratively toward the same set of goals.

T-BARS is best suited for coding verbal protocol data collected during the conduct of tactical exercises. Verbal protocols can produce a rich source of information about how the learner is thinking through the tactical problem, and about the rationale behind his or her actions and judgments. The *T-BARS User Guide* suggests protocols to employ to produce data that is most revealing of the learner's cognition. We believe the tool is also amenable for coding written measures of performance produced from a training session or for conducting ratings during live observations. With regard to coding written passages of text, T-BARS is probably most useful when the user has input into the queries and probes presented to the tactician. The goal should be to capture not only the decisions or orders, but also the learner's interpretation of the situation and rationale for the actions. To rate performance during live tactical exercises, the user of T-BARS should be very familiar with the assessment tool and its content. The mental workload for the rater will be high as exercises tend to progress quickly and tacticians can discuss several concepts in a short span of time. In the *T-BARS User Guide* we recommend approaches to data collection during live observations that minimize workload to the greatest extent possible.

Technology Evaluation. As part of the development cycle for advanced battle command technologies, one of the questions to address is the influence of the technology on user cognition. With T-BARS, we have a tool for measuring whether battle command tools enable tactical decision makers to function at higher levels of cognitive proficiency than they would otherwise. Recall that the aspects of tactical thinking are cognitive processes rather than mental models – considering timing, seeing the big picture, remaining flexible and thinking about contingencies, and visualizing the battlefield – develop later in an individual's career as experience is gained. Within T-BARS, these are represented by and large at levels 4 and 5. These are the cognitive manipulations that advanced battle command technologies typically aim to support. As an example, some visualization technologies purport to give the commander a better view of the entire battlefield, on dimensions of time and space, whereby he can intuitively understand the current situation and better predict the impact of future candidate actions. If indeed a visualization tool enables better prediction of the consequences of actions, we should see commanders achieving higher ratings on the T-BARS scales – 4's and 5's – with the technology than without it.

One danger of using advanced technologies is that they can actually hinder rather than support the user's cognitive processes (Crandall, Klein, & Hoffman, in preparation; Klein, 2000).

This is especially true for individuals who are already operating at very high levels of cognitive proficiency with rich and finely discriminated mental models. For example, some technologies intended to support weather forecasting capabilities have reportedly resulted in decreased accuracy for expert forecasters (Crandall et al., in preparation; Klein, 2000). These tools take large amounts of data and produce smooth curves and general trends for the forecaster. However, experts have learned to look for jaggedness in the data representing pockets of discrepant activity to predict how various forces will interact to produce what we experience as "weather." The technologies smooth the jagged edges and thereby take away a significant part of the weather picture for the experts. In this way, experts are less effective using the technologies than without them. Likewise, it is necessary to ensure that battle command technologies do not cripple tactical experts in the same ways, by taking away indicators that stand out from the rest of the data but actually represent an important situational aspect. By using T-BARS to measure tactical performance with and without technological support, it is possible to ensure that we are not implementing tools that bring level 5 tacticians down to 3's or 4's.

Future Directions

This effort has produced a reliable tool for assessing tactical thinking mental models. The next step in the development of the T-BARS is to establish the validity of its scales to ensure that it indeed measures mental models as intended. In addition, there is a need to collect usability feedback and/or data from other users of the T-BARS to ensure that its application is well understood and generally consistent across researchers. If this is to be an assessment tool that is widely used to evaluate training and technological interventions, it is critical that researchers are employing it in similar ways across the range of assessment settings to facilitate comparisons of findings. We are therefore interested in establishing a community of practice in the short term to collect input regarding how various researchers are applying the tool and with what types of results.

We believe that a BARS approach to measuring mental models and thus cognitive proficiency can be more broadly applied within the military. There is an opportunity to produce BARS for other sub-domains such as Intelligence or Information Operations. It may even be possible to develop BARS for team mental models that could be applied broadly to examine group collaboration and functioning regardless of the specific context of type of team. The theoretical foundation for other BARS within military specialty areas has been established by mapping the levels of performance to general cognitive profiles as described by the Dreyfus and Dreyfus (1986) stage model of cognitive skill acquisition. We believe the process employed to develop T-BARS – iterative generation and testing of behavioral descriptors within each of the five levels – was effective and can be used in future related efforts.

References

Baxter, H. C., Harris-Thompson, D., & Phillips, J. K. (2004, December 6-9, 2004). *Sensemaking: cognitive approach to training situation awareness.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Benner, P. (1984). *From novice to expert: Excellence and power in clinical nursing practice.* Reading, MA: Addison-Wesley.

Benner, P. (2004). Using the Dreyfus model of skill acquisition to describe and interpret skill acquisition and clinical judgment in nursing practice and education. *Bulletin of Science, Technology & Society, 24*(3), 189-199.

Bloom, B.C. (Ed.). (1956). *Taxonomy of educational objectives: Handbook i, cognitive domain.* New York; Toronto: Longmans, Green: David McKay Company, Inc.

Chi, M. (1978). Knowledge structures and memory development. In R. Siegler (Ed), *Children's thinking: What develops?* (pp. 73-96). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1980). *Representation of physics knowledge by experts and novices (technical report no. 2)* (No. Technical Report No. 2). Pittsburgh, PA: University of Pittsburgh: Learning Research and Development Center.

Crandall, B., Klein, G., & Hoffman, R. R. (in preparation). *Working minds: A practitioner's guide to cognitive task analysis.* Cambridge, MA: The MIT Press.

Deckert, J. C., Entin, E. B., Entin, E. E., MacMillan, J., & Serfaty, D. (1996). *Military decisionmaking expertise* (ARI Research Note 96-15). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuitive expertise in the era of the computer.* New York: The Free Press.

Einhorn, H. J. (1980). Learning from experience and suboptimal rules in decision making. In T. S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 1-20). Mahwah, NJ: Lawrence Erlbaum & Associates.

Feltovich, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). Lcs: The role and development of medical knowledge in diagnostic expertise. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence: The first decade* (pp. 275-319). Reading: Addison-Wesley.

Gentner, D. (2002). Mental models, psychology of. In N. J. Smelser & P. B. Bates (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 9683-9687). Amsterdam, The Netherlands: Elsevier Science.

Glaser, R. (1996). Changing the agency for learning: Acquiring expert performance. In K.A. Ericsson (Ed.), *The Road to Excellence* (pp.1-50). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Glaser, R., & Baxter, G. P. (2000). *Assessing active knowledge* (Technical Report for Center for the Study of Evaluation). Los Angeles, CA: University of California.

Houldsworth, B., O'Brien, J., Butler, J., & Edwards, J. (1997). Learning in the restructured workplace: A case study. *Education and Training, 39*(6), 211-218.

Klein, G. (1993). Twenty questions: Suggestions for research in naturalistic decision making. In G. Klein, J. Orasanu, R. Calderwood, C. Zsambok, (Eds), *Decision making in action: Models and methods* (pp. 389-403). Westport, CT, US: Ablex Publishing.

Klein, G. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Klein, G. (1999). Applied decision making. In P. Hancock (Ed.), H*uman performance and ergonomics* (pp. 87-107). San Diego, CA, US: Academic Press, Inc.

Klein, G. (2000). Can information technology reduce expertise? In *Proceedings of the human performance, situation awareness and automation: User-centered design for the new millennium*. Savannah, GA: Omnipress CD-Rom.

Klein, G., & Crandall, B. (1995). The role of mental simulation in problem solving and decision making. In P. Hancock, J. Flach, J. Caird, K. Vicente, (Eds), *Local applications of the ecological approach to human-machine systems* (pp. 324-358). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Klein, G., & Hoffman, R. (1993). Seeing the invisible: Perceptual/cognitive aspects of expertise. In M. Rabinowitz (Ed.), *Cognitive science foundations of instruction* (pp. 203-226). Mahwah, NJ: Lawrence Erlbaum & Associates.

Klein, G., Phillips, J. K., Battaglia, D. A., Wiggins, S. L., & Ross, K. G. (2002). *Focus: A model of sensemaking. Interim report - year 1 prepared under contract 1435-01-01-ct-31161 [Dept. of the Interior] for the U.S. Army research institute for the behavioral and social sciences, Alexandria, VA* (Interim Report - Year 1 prepared under Contract 1435-01-01-CT-31161 [Dept. of the Interior] for the U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA). Fairborn, OH: Klein Associates Inc.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *IEEE Intelligent Systems, 18*(3), 81-85.

Klein, G., & Wolf, S. (1998). The role of leverage points in option generation. *IEEE Transactions on Systems, Man and Cybernetics: Applications and Reviews, 28*(1), 157-160.

Larkin, J. H. (1983). The role of problem representation in physics. In D. Gentner & A. L. Stevens (Eds.), *Mental models*. Mahwah, NJ: Lawrence Erlbaum & Associates.

Lipshitz, R., & Ben Shaul, O. (1997). Schemata and mental models in recognition-primed decision making. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 293-304). Mahwah, NJ: Lawrence Erlbaum & Associates.

Lussier, J. W. (1998). *Developing expertise in battlefield thinking: Models of battlefield thinking* (Unpublished briefing). Ft. Leavenworth, KS: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lussier, J. W., Ross, K. G., & Mayes, R. (2000). *Coaching techniques for adaptive thinking.* Paper presented at the 2000 Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Lussier, J. W., Shadrick, S. B., & Prevou, M. I. (2003). *Think like a commander prototype: Instructor's guide to adaptive thinking* (ARI Research Product 2003-01). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

McElroy, E., Greiner, D., & de Chesnay, M. (1991). Application of the skill acquisition model to the teaching of psychotherapy. *Archives of Psychiatric Nursing, 5*(2), 113-117.

Militello, L. G., & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, Special Issue: Task Analysis, 41*(11), 1618-1641.

Muchinsky, P. M. (2003). *Psychology applied to work.* Melmont, CA: Wadsworth/Thomson Learning.

Phillips, J., Klein, G., & Sieck, W. (2004). Expertise in judgment and decision making: A case for training intuitive decision skills. In D. Koehler, N. Harvey (Eds), *Blackwell handbook of judgment and decision making* (pp. 297-315). Malden, MA, US: Blackwell Publishing.

Phillips, J. K., Shafer, J., Ross, K. G., Baxter, H., & Harris, D. S. (2003). *Framework for assessment of tactical decision-making simulations: Results and discussion for final report* (Final Report prepared under Subcontract CHI 00001-07 (Prime IDIQ N61339-99-D0012) for CHI Systems). Fairborn, OH: Klein Associates Inc.

Phillips, J. K., Ross, K. G., & Shadrick, S. B. (in preparation). User's guide for Tactical Thinking Behaviorally Anchored Rating Scales (ARI Research Product). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Ross, K. G., Battaglia, D. A., Hutton, R. J. B., & Crandall, B. (2003). *Development of an instructional model for tutoring tactical thinking* (Final Technical Report for Subcontract No. SHAI-COMM-01; Prime Contract No. DASW01-01-C-0039 submitted to Stottler Henke Associates Inc., San Mateo, CA). Fairborn, OH: Klein Associates.

Ross, K. G., Battaglia, D. A., Phillips, J. K., Domeshek, E. A., & Lussier, J. W. (2003). Mental models underlying tactical thinking skills (submitted). In *Proceedings of the Interservice/industry training, simulation, and education conference 2003.*

Ross, K. G., & Lussier, J. W. (1999). *A training solution for adaptive battlefield performance.* Paper presented at the 1999 Interservice/Industry Training, Simulation, and Education Conference, Orlando, FL.

Ross, K. G., Phillips, J. K., Klein, G., & Cohn, J. (2005). *Creating expertise: A framework to guide technology-based training* (Final Report prepared under Contract M67854-04-C-8035 for Marine Corps System Command/Program Manager for Training Systems under the sponsorship of the Office of Naval Research MARCORSYSCOM/PMTRASYS, Orlando, FL). Fairborn, OH: Klein Associates Inc.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits on the search for mental models. *Psychological Bulletin, 100*(3), 349-363.

Serfaty, D., MacMillan, J., Entin, E. E., & Entin, E. B. (1997). The decision-making expertise of battle commanders. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 233-246). Mahwah, NJ: Lawrence Erlbaum & Associates.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252-266.

Simon, H. A. (1975). The functional equivalence of problem solving skills. Psychology, 7, 268-288.

Stemler, S. (2004). Automated essay scoring: A human's review. *PsycCritiques, 49.*

Appendix A

Cognitive Profiles from the Dreyfus and Dreyfus (1986) Stage Model of Cognitive Skill Acquisition

STAGE 1: NOVICE	
General Characteristics	
Knowledge	***Performance***
•Objective facts and features of the domain (Dreyfus & Dreyfus, 1986). •Context-free (abstract) rules to guide behavior (Dreyfus & Dreyfus, 1986). •Domain characteristics acquired through textbooks and classroom instruction (Benner, 1984).	•Guided by rules; is limited and inflexible (Benner, 1984). •Shows recognition of elements of the situation without considering context (Dreyfus & Dreyfus, 1986). •Is variable and awkward (Glaser, 1996). •Focuses on isolated variables (Glaser, 1996). •Consists of a set of individual acts rather than an integrated strategy (Glaser, 1996; McElroy et al., 1991). •Is self-assessed based on how well he adheres to learned rules (Benner, 1984; Dreyfus & Dreyfus, 1986). •Reflects a sense of being overwhelmed since all stimuli are perceived to be equally relevant (McElroy et al., 1991).

STAGE 2: ADVANCED BEGINNER	
General Characteristics	
Knowledge	***Performance***
•Some domain experience (Benner, 1984; Dreyfus & Dreyfus, 1986). •More objective, context-free facts than the novice, and more sophisticated rules (Dreyfus & Dreyfus, 1986). •Situational elements, which are recurring, meaningful elements of a situation based on prior experience (Dreyfus & Dreyfus, 1986). •A set of self-generated guidelines that dictate behavior in the domain (Benner, 1984). •Seeks guidance on task performance from context-rich sources (e.g., experienced people, documentation of past situations) rather than rule bases (e.g., textbooks) (Houldsworth et al., 1997).	•Is marginally acceptable (Benner, 1984). •Combines the use of objective, or context-free, facts with situational elements (Dreyfus & Dreyfus, 1986). •Ignores the differential importance of aspects of the situation; situation is a myriad of competing tasks, all with same priority (Benner, 1984; Dreyfus & Dreyfus, 1986; Shanteau, 1992). •Shows initial signs of being able to perceive meaningful patterns of information in the operational environment (Benner, 1984). •Reflects attitude that answers are to be found from an external source (Houldsworth et al., 1997). •Reflects a lack of commitment or sense of involvement (McElroy et al., 1991).

STAGE 3: COMPETENT	
General Characteristics	
Knowledge	***Performance***
•How to think about the situation in terms of overarching goals or tasks (Benner, 1984). •The relative importance of subtasks depending on situational demands (Benner, 1984; Dreyfus & Dreyfus, 1986). •Particular patterns of cues suggest particular conclusions, decisions, or expectations (Dreyfus & Dreyfus, 1986). •A personalized set of guiding principles based on experience (Houldsworth et al., 1997). •How to anticipate future problems (Houldsworth et al., 1997).	•Is analytic, conscious, and deliberate (Benner, 1984; Dreyfus & Dreyfus, 1986). •Does not rely on a set of rules (Houldsworth et al., 1997). •Is efficient and organized (Benner, 1984; Dreyfus & Dreyfus, 1986). •Is driven by an organizing plan that is generated at the outset of the situation (Dreyfus & Dreyfus, 1986). •Reflects an inability to digress from the plan, even when faced with new, conflicting information (Dreyfus & Dreyfus, 1986). •Reflects an inability to see newly relevant cues due to the organizing plan or structure that directs attention (Benner, 2004). •Reflects an emotionally involved performer who takes ownership of successes and failures (Dreyfus & Dreyfus, 1986). •Focuses on independent features of the situation rather than a synthesis of the whole (Houldsworth et al., 1997).

STAGE 4: PROFICIENT	
General Characteristics	
Knowledge	***Performance***
•Typical "scripts" for categories of situations (Klein, 1998). •How to set expectancies and notice when they are violated (Benner, 1984). •How to spot the most salient aspects of the situation (Benner, 1984; Dreyfus & Dreyfus, 1986). •Personalized maxims, or nuances of situations, that require a different approach depending on the specific situation, but not how to apply the maxims correctly (Benner, 1984; Houldsworth et al., 1997).	•Reflects a perception of the situation as a whole rather than its component features (Benner, 1984). •Is quick and flexible (Benner, 1984). •Reflects a focus on long-term goals and objectives for the situation (Benner, 1984). •Utilizes prior experience (or intuition) to assess the situation, but analysis and deliberation to determine a course of action (Dreyfus & Dreyfus, 1986; McElroy et al., 1991). •Reflects a synthesis of the meaning of information over time (Benner, 2004). •Reflects a more refined sense of timing (Benner, 2004).

STAGE 5: EXPERT	
General Characteristics	
Knowledge	***Performance***
•How to make fine discriminations between similar environmental cues (Klein, 1993). •How to intuitively assess the situation (Benner, 2004; Dreyfus & Dreyfus, 1986). •How to respond to maxims or nuances based on the unique array of cues and factors in the situation (Benner, 2004). •How to intuitively respond to the situation (Benner, 1984; Dreyfus & Dreyfus, 1986). •How tasks and subtasks are supposed to be performed (Phillips, Klein, & Sieck, 2004). •How equipment and resources function in the domain (Phillips et al., 2004). •How to perceive meaningful patterns in large and complex sets of information (Klein, 1998; Dreyfus & Dreyfus, 1986). •What is typical and atypical for a particular situation (Dreyfus & Dreyfus, 1986; Feltovich, Johnson, Moller, & Swanson, 1984; Klein, 1999). •A wide range of routines or tactics for getting things done (Klein, 1999). •More facts about the domain than less proficient individuals (Phillips et al., 2004). •A huge library of lived distinguishable experiences that impact handling of new situations (Dreyfus & Dreyfus, 1986). •How to set expectancies and notice when they are violated (Benner, 1984).	•Is fluid and seamless, like walking or talking; "integrated rapid response" (Benner, 1984, 2004; Dreyfus & Dreyfus, 1986). •Is based on prior experience for both assessment and decision making (Dreyfus & Dreyfus, 1986). • Is automatic, and the rationale for actions is often difficult to articulate (Benner, 1984). •Relies heavily and successfully on mental simulation to predict events, diagnose prior occurrences, and assess courses of action (Einhorn, 1980; Klein & Crandall, 1995). •Consists of more time assessing the situation and less time deliberating a course of action (Lipshitz & Ben Shaul, 1997). •Shows an ability to detect problems and spot anomalies early (Feltovich et al., 1984). •Capitalizes on leverage points, or unique ways of utilizing ordinary resources (Klein & Wolf, 1998). •Reflects use of innovations and new possibilities for responding to particular situations (like leverage points) (Benner, 2004). •Manages uncertainty with relative ease, by filling gaps with rational assumptions and formulating information-seeking strategies (Klein, 1998; Serfaty, MacMillan, Entin, & Entin, 1997). •Reflects metacognitive skill, or the ability to self-monitor (Chi, 1978; Chi, Feltovich, & Glaser, 1980; Larkin, 1983; Simon, 1975). •Shows efficient information search activities (Shanteau, 1992).

Appendix B

Tactical Thinking Profiles for Each T-BARS Theme

Theme 1. Know and Use All Assets Available. Combat leaders must not lose sight of the synergistic effects of fighting their command as a combined arms team - this includes not only all assets under their command, but also those which higher headquarters might bring to bear to assist them.

Knows Textbook Capabilities	Matches Assets to Mission Requirements	Utilizes Organic Assets to Accomplish Mission Objectives	Recognizes Full Range of Assets Required based on Situational Demands	Applies Full Range of Assets to Direct the Outcome of the Battle
1	2	3	4	5
Performance is abstract and rule-based, and focuses on variables in isolation. Individual knows facts about standard capabilities of organic assets such as ranges of weapons, number of vehicles per unit, and so forth. The foundational knowledge required to analyze how assets can be applied to the situation has not yet developed.	*Performance reflects simple analytical processing using a limited experience base. Organic assets are matched to mission requirements. For example, a tank formation would be allocated to the area where heavy armor is needed for protection. Individual has difficulty prioritizing tasks, so asset utilization is driven by capabilities (what the asset can do) over situational demand (what is the most pressing mission task).*	*Performance reflects a mental model of asset utilization, but remains dependent on analysis and planning rather than recognition and intuition. Individual can prioritize mission tasks and predict how the situation could unfold, and an asset utilization plan is generated against that analysis. However, execution is driven by the plan over the situation, so individual has difficulty adjusting asset utilization to meet changing situational demands.*	*Performance reflects a recognitional or intuitive assessment of the situation, but analytical decision making where the individual deliberates about a course of action. Individual recognizes the availability of non-organic and non-military assets in addition his own organic assets. For example, civilians are recognized to be valuable sources of HUMINT. Situational demands drive asset utilization, rather than the plan or the organic assets at the individual's disposal.*	*Performance reflects a recognitional ability to assess and decide. Individual can visualize specific outcomes of asset utilization and has the ability to avoid unwanted consequences. For example, he knows how to command and maneuver his forces to avoid an uprising by the locals. Individual leverages and coordinates organic, non-organic, and non-military assets to achieve mission objectives.*

Theme 2. Keep a Focus on the Mission and Higher's Intent. Combat leaders must never lose sight of the purpose and results they are directed to achieve - even when unusual and critical events may draw them in a different direction.

Focuses on Own Mission	Discriminates Intent and Explicit Mission	Models Effects of Own Mission and HQ Intent	Makes Accurate Predictions	Supports Intent
1	2	3	4	5
Performance is abstract and rule-based, and focuses on variables in isolation. Individual fixates on own mission rather than considering larger organization's mission. He is unable to consider higher intent. The foundational knowledge required to analyze steps required for mission accomplishment has not yet developed.	*Performance reflects simple analytical processing using a limited experience base. Mission tasks are paramount to all else, and intent can be articulated but not operationalized. Individual has difficulty prioritizing tasks for mission accomplishment, and is often uncertain or overwhelmed as situation evolves. There is a tendency to rely on direction from higher HQ rather than making own decisions.*	*Performance reflects a mental model of how intent is achieved through mission tasks, but remains dependent on analysis and planning rather than recognition and intuition. Individual can prioritize mission tasks and predict how the situation could unfold, and a course of action is generated based on that analysis. However, performance in execution is guided by an efficient but rigid plan that is not adapted to account for changes in the situation.*	*Performance reflects a recognitional or intuitive assessment of the situation, but analytical decision making where the individual deliberates about a course of action. Individual recognizes how situational factors impact the mission and the path to achieving intent. For example, he can visualize the enemy's likely objective and use of terrain, and he uses that assessment to deliberate about how to support the intent through his own mission. During execution, changes in the situation are recognized intuitively and mission tasks are adapted or changed to continue to support intent.*	*Performance reflects a recognitional ability to assess and decide. Individual can quickly and accurately assess the situation, visualize contingencies, and devise an action plan that accomplishes the intent while avoiding unwanted 2nd and 3rd order consequences. Individual operates from a big picture perspective in which he takes actions that support the short- and long-term objectives of the coalition force.*

Theme 3. Model a Thinking Enemy or Populace. Combat leaders must not forget that the adversary is a reasoning human being, intent on defeating them - it's tempting to simplify the battlefield by treating the enemy as static or simply reactive. Likewise, the local populace has its own motivations that drive its actions within the battlespace.

Uses Enemy Templates	Regards Enemy as Static	Regards Enemy as Intelligent and Dynamic	Predicts Enemy Actions	Denies Enemy Intent
1	2	3	4	5
Performance is abstract and rule-based. Individual acknowledges enemy superficially and equates him with theoretical or doctrinal templates learned in schoolhouse – for example, the typical Soviet offensive formation. The foundational knowledge required to analyze probable enemy actions and objectives has not yet developed.	*Performance reflects simple analytical processing using a limited experience base. Enemy is understood to have an impact on the mission, but is regarded as a static, non-thinking adversary. Individual has trouble distinguishing enemy centers of gravity from the rest of the enemy picture. Individual struggles to make sense of or draw hypotheses about the enemy's objectives.*	*Performance reflects a mental model of an intelligent, dynamic enemy. Individual analyzes the enemy situation and predicts enemy actions in order to formulate an efficient and organized COA that defeats the enemy. Ideas about the enemy's objectives and COA are constructed, but they are general and imprecise. Because the individual is guided by the plan rather than situational demands, he struggles to adapt his COA when the enemy situation changes during execution.*	*Performance reflects a recognitional or intuitive assessment of the enemy's objective and intent, but analytical decision making where the individual deliberates about a course of action that defeats the enemy. Individual continually updates his assessment of the enemy situation and his predictions about the enemy's next steps based on situational factors.*	*Performance reflects a recognitional ability to assess and decide. Individual visualizes how enemy will act and react, and takes actions to deny enemy intent. For example, he recognizes enemy leverage points and takes action to neutralize them or make them unavailable.*

Theme 4. Consider Effects of Terrain. Combat leaders must not lose sight of the operational effects of the terrain on which they must fight - every combination of terrain and weather has a significant effect on what can and should be done to accomplish the mission.

Uses Terrain Checklists	Identifies Important Terrain Features	Incorporates Terrain into Own Plan	Recognizes How the Enemy May Use Terrain	Turns Terrain to Own Advantage
1	2	3	4	5
Performance focuses on identifying discrete features of terrain. Individual uses standard checklists to determine relevant terrain features. The foundational knowledge required to analyze the impact of terrain on the mission has not yet developed.	*Performance reflects simple analytical processing using a limited experience base. Important terrain features are identified and prominent problem areas such as chokepoints are avoided. However, individual remains unable to leverage terrain to own advantage.*	*Performance reflects a mental model of the impact of terrain on the mission. Individual performs an analysis of the terrain and incorporates terrain features into the plan. For example, in an urban setting the tallest and sturdiest buildings are perceived as good locations to occupy. However, the individual tends to adhere to the plan even after the situation has evolved and new information about the terrain becomes available.*	*Performance reflects a recognitional or intuitive assessment of the aspects and patterns of terrain that are critical for friendly and enemy forces, but deliberate analysis of how to utilize the terrain to accomplish the mission. Individual continually updates his view of the terrain and its impact on the mission as the situation evolves and new terrain features and patterns are discovered.*	*Performance reflects a recognitional ability to assess and decide. Individual is quickly able to visualize how terrain will impact the friendly mission and predicted enemy actions. He leverages the terrain to his own advantage and denies the enemy's ability to do the same.*

Appendix C

Final Version of T-BARS

Theme 1. Know and Use All Assets Available. Combat leaders must not lose sight of the synergistic effects of fighting their command as a combined arms team - this includes not only all assets under their command, but also those which higher headquarters might bring to bear to assist them.

1	2	3	4	5
(A) Asks questions about facts of own organic assets.	(A) Identifies how assets can be used in a general sense (e.g., unmanned aerial vehicles (UAVs) can be used for recon), but not how to maximize for	(A) Articulates how specific organic assets can be used to overcome enemy capabilities and accomplish the mission.	(A) Articulates rationale for employing a particular organic asset based on situational factors.	(A) Leverages non-organic assets from larger organization.
(B) States facts about what assets are organic to own unit.	(B) Makes a straight match of organic asset(s) to portion(s) of the mission without regard to prioritization of effort.	(B) Identifies trade-offs, benefits and risk of splitting or reassigning assets.	(B) Makes a statement about the availability and/or value of non-organic assets.	(B) Articulates how non-organic assets can be accessed.
(C) States facts about capabilities of organic assets.	(C) Describes general posture for organic assets to take rather than specific tasks.	(C) Articulates rationale for use of specific assets for particular task or mission (e.g., armored vehicles needed for safety).	(C) Makes statements about own and other units as a team rather than isolated entities.	(C) Assembles assets in an integrated fashion based on rapid assessment of situation.
(D) Gives "templated" answers about how assets will be used/restates mission information.	(D) Questions whether assets (e.g., size of force) are adequate for mission or contingencies.	(D) Describes or makes reference to trade-offs of employing assets or keeping them in reserve.	(D) Makes a statement about the availability and/or importance of non-military assets such as civilians.	(D) Makes a statement about assets in terms of what other units need. [Big Picture]

1	2	3	4	5
(E) States facts about status of own assets during execution.	(E) Makes simple acknowledgement that timing of asset utilization is a consideration. [Timing]	(E) Mixes and matches organic assets/units for a common purpose.	(E) Changes plans for assets/asset usage during execution when initial plan is not working.	(E) Makes reference to time needed for other units to act based on their assets. [Big Picture]
(F) Asks questions about status of own assets during execution.	(F) Articulates a consideration of the safety and security of assets (including Soldiers).	(F) Articulates size of force needed for particular mission or task.	(F) Provides rationale for changing the utilization of assets.	(F) Communicates with other units to coordinate action. [Big Picture]
(G) Provides facts about status of own assets during execution.	(G) Communicates current situation to subordinate units during execution.	(G) Employs organic assets to proactively acquire information.	(G) Mixes and matches organic assets based on situational demands (e.g., combines engineer assets or MPs with infantry unit).	(G) Articulates plan for asset usage based on primary effect as well and second and third order effects and consequences. [Visualization]
(H) Describes use of assets without regard to timing and/or terrain.		(H) Calculates time distance based on knowledge of terrain, obstacles, weather, asset capability, etc. [Timing]	(H) Describes asset utilization in terms of meeting time constraints. [Timing]	(H) Refines plan for asset usage during execution based on primary effect as well and second and third order effects and consequences. [Visualization]

1	2	3	4	5
		(I) Indicates the impact of timing and/or sequencing of asset use (e.g., Armored Vehicle Launched Bridge (AVLB) is a very slow tracked vehicle, so you have to anticipate the time it takes to transport it.) [Timing]	(I) Discriminates relative strengths and weaknesses of assets (organic, non-organic, or non-military] within the context of a mission.	(I) Articulates specific plans for use of non-military assets.
		(J) Describes unfolding of events or sequencing as assets are employed. [Visualization, Timing]	(J) Identifies danger area(s) for asset(s) based on terrain and/or potential enemy location or action. [Terrain, Enemy, Visualization]	(J) Articulates how limited assets will be used to produce large effects.
		(K) Matches specific asset to particular location and time. [Visualization]	(K) Discusses sustainability of organic assets (e.g., with food, water, ammo, etc.) in light of situational demands.	(K) Identifies and uses emergent leverage points during execution to maximize asset effects.
		(L) Articulates first order effects of using specific assets at specific times. [Big Picture]	(L) States needs to conserve resources.	

1	2	3	4	5
		(M) Predicts consequences of using assets or not using assets.	(M) Deploys assets in an integrated fashion (not as isolated systems) to achieve larger tactical purpose.	
		(N) Projects what other assets might be needed or useful. [Visualization]	(N) States rationale for asset usage in terms of retaining flexibility of usage. [Visualization, Contingency Thinking]	
		(O) Discusses what assets might be useful for potential specific contingencies. [Contingency Thinking]	(O) States consequences or effects of asset usage beyond specific primary effect. (Second and third order effects).	
		(P) Identifies approach for using asset(s) for particular contingency(s). [Contingency Thinking]	(P) Articulates rationale for timing and/or sequencing of asset usage. [Timing]	
		(Q) Asks questions about availability of non-organic assets. [Big Picture Thinking]	(Q) Articulates constraints due to troop availability vis-à-vis mission.	

Theme 2. Keep a Focus on the Mission and Higher's Intent. Combat leaders must never lose sight of the purpose and results they are directed to achieve - even when unusual and critical events may draw them in a different direction.

1	2	3	4	5
(A) Asks questions about facts of scenario description.	(A) Makes a statement about the situation (planning and execution) in terms of mission analysis without intent as a "lens."	(A) Analyzes intent statement in order to determine what has to be accomplished.	(A) Articulates rationale for sequencing tasks based on situational factors.	(A) States concept of operations rapidly after receipt of mission, citing variable of situation and/or intent.
(B) Asks questions about facts of stated mission tasks.	(B) Exhibits uncertainty about priorities in the mission or does not prioritize.	(B) Debates whether mission will support intent.	(B) Describes how own mission will contribute to larger operation/mission.	(B) Articulates how and/or why course of action or concept of operations supports higher intent.
(C) Articulates understanding of mission without regard to intent.	(C) Expresses uncertainty about what constitutes mission success.	(C) Identifies consequences of failing to complete mission in terms of effect on intent/higher operations. [Big Picture]	(C) Discusses during execution whether actions are supporting intent.	(C) Identifies changes or relevant new information in situation and articulates adjustments to course of action during execution.

1	2	3	4	5
(D) Asks for clarification of rules of engagement (ROE).	(D) States what needs to be accomplished (mission task) but not how to do it.	(D) Identifies the need to prioritize mission tasks or subtasks.	(D) Prioritizes actions in order to support higher intent or larger operations (e.g., "I need to do 'this' instead of 'that.' ")	(D) Allocates assets during planning or execution based on a prediction about the enemy.
(E) Uses "templated" methods for analyzing/planning.	(E) Identifies timing as a consideration in mission tasks. [Timing]	(E) Differentiates priorities in mission tasks.	(E) Predicts how future events can impact own mission and generates contingencies to overcome interference.	(E) Proactively places assets to support larger intent. [Contingency Thinking]
(F) Uses only organic assets for mission.	(F) Identifies information requirements that can impact mission (e.g., what are dangers ahead on route).	(F) Articulates what would be a favorable outcome for a particular task.	(F) Articulates during execution changes that will interfere with achieving intent.	(F) Responds to change fluidly by implementing planned contingency or rapidly articulating new contingency. [Contingency Thinking, Visualization]
(G) Adheres rigidly to stated tasks of mission.	(G) Keeps Higher HQ informed about plans and situation during execution.	(G) Describes future events that may impact or interfere with current mission.	(G) Describes potential impact of non-combatant activity on mission during execution.	(G) Eliminates obstacle to higher intent.

1	2	3	4	5
(H) Asks about facts of events during execution.	(H) Relies on Higher HQ to make decisions.	(H) Articulates what task will happen next.	(H) Describes how situation could draw unit away from mission accomplishment (i.e., mission creep).	(H) Creates advantage for higher or adjacent unit. [Big Picture]
(I) Communicates only within unit about plans and situation.		(I) Articulates in a general manner (i.e., without detail) what the effect(s) of task/ mission accomplish-ment will be beyond intended effect.	(I) Requests additional support from higher when mission accomplishment requires it.	(I) Presents opportunities to higher or adjacent units. [Big Picture]
(J) Asks questions about scenario facts or events during execution.		(J) States a consideration of actions other than stated mission (implied tasks or additional tasks to meet intent).	(J) Articulates consideration of two or more of the following together: Mission statement; own tasks; higher intent; desired end state; what success looks like.	(J) Articulates specific second and third order effects of an action during execution.
(K) Questions own decision making authority.		(K) Seeks clarification of implied tasks.	(K) Alters or refines course of action in light of changes in situation.	(K) Articulates actions necessary to ensure mission accomplishment when faced with threat to mission success during execution.

1	2	3	4	5
		(L) Seeks clarification on potential "be prepared" missions.	(L) Infers priorities from intent statement.	(L) Articulates actions necessary to ensure intent when stated mission is superseded by dynamic events, but intent is still achievable.
		(M) Adheres to plan (tasks and manner of accomplishment) in the face of new and changing information.	(M) Alters or refines course of action based on discovery of new leverage points in situation.	
		(N) Forges ahead with course of action when it is no longer relevant.	(N) Articulates how events will be sequenced.	
		(O) Identifies important aspects of mission that require attention for success. [Visualization]	(O) Articulates how timing or sequencing of events needs to allow for replanning or contingencies.	

1	2	3	4	5
		(P) States general sequencing of tasks without articulating how to accomplish each task.		
		(Q) Articulates timing estimates or sequencing as a critical component of the planning process. [Timing]		
		(R) Notes the importance of timing and/or sequencing for the particular mission. [Timing]		

Theme 3. Model a Thinking Enemy or Populace. Combat leaders must not forget that the adversary is a reasoning human being, intent on defeating them - it's tempting to simplify the battlefield by treating the enemy as static or simply reactive. Likewise, the local populace has its own motivations that drive its actions within the battlespace.

1	2	3	4	5
(A) Ignores enemy during mission analysis/planning/ execution.	(A) Articulates enemy capabilities with limited or no consideration of current situation's context.	(A) Generates ideas about what enemy might be thinking.	(A) Articulates an assessment of enemy objective, approach, or size/strength based on situational factors.	(A) Develops a rationale as the basis to deny the enemy intent.
(B) Ignores typical enemy capabilities & assets or states them incorrectly.	(B) Makes a general statement about the enemy's approach.	(B) Generates ideas about what enemy's objective might be.	(B) Articulates that own course of action should deny enemy intent, but is unsure how to operationalize.	(B) Articulates how course of action will use terrain, assets, or other resources to deny enemy objective.
(C) Ignores typical enemy tactics or states them incorrectly.	(C) Expresses general concern about enemy situation.	(C) Generates hypotheses about how the enemy might carry out a course of action. [Visualization]	(C) Evaluates two or more courses of action to determine which better inhibits enemy.	(C) Makes a projection about how enemy or populace will react to own actions.
(D) Gives "templated" answers about expected enemy actions (e.g., "typical enemy will...")	(D) Identifies enemy capabilities with regard to impact on own mission.	(D) Questions how enemy might respond to own COA.	(D) Articulates an assessment of how enemy will use the terrain, employ their assets, or use avenues of approach.	(D) Articulates action plan during execution to counter suspected enemy intent.

1	2	3	4	5
(E) Describes enemy as a single, general entity rather than as subsystems or several smaller units.	(E) Speculates about potential enemy actions.	(E) Articulates enemy's capabilities and/or vulnerabilities in light of terrain or other situational factors. [Visualization]	(E) Identifies danger area for own forces and/or leverage points for enemy based on situational factors.	(E) Articulates assessment of enemy operation that has been conducted or is being conducted.
(F) Asks for or states facts about enemy actions during execution (no interpretation).	(F) Generates hypotheses about who may be an enemy/threat during execution.	(F) States expectations about enemy activity based on patterns or specific elements in current situation.	(F) Identifies specific piece of desired information about enemy/populace.	(F) Generates hypotheses about who may be an enemy/threat during execution.
(G) Puts a blue unit in a dangerous spot vis-à-vis enemy.	(G) States only what is observed about enemy during execution (what he is doing) without inference.	(G) Generates ideas about how enemy might use terrain, employ assets, or use avenues of approach. [Visualization]	(G) Generates ideas about what enemy might be doing during execution.	
	(H) Asks questions/seeks information about what enemy is doing in own sector.	(H) Asks questions/seeks information about what enemy is doing in another sector.	(H) Generates ideas during execution about what enemy will do next. [Visualization]	
		(I) Articulates a general expectation about enemy activities or responses to friendly actions (e.g., "I expect light resistance.")	(I) Discusses varied responses/actions to take in response to potential enemy actions. [Contingency Thinking]	

1	2	3	4	5
		(J) Articulates disadvantages the enemy's action has created. [Big Picture]		
		(K) Articulates specific contingencies (while planning) in response to enemy actions that might be taken. [Contingency Thinking, Visualization]		
		(L) Adheres to execution path in the face of new or changing information about enemy.		

Theme 4. Consider Effects of Terrain. Combat leaders must not lose sight of the operational effects of the terrain on which they must fight - every combination of terrain and weather has a significant effect on what can and should be done to accomplish the mission.

1	2	3	4	5
(A) Uses checklist to assess terrain (e.g., OCOKA).	(A) Asks basic questions about terrain features.	(A) Appraises, during planning, how individual terrain features are likely to impact own mission.	(A) Articulates how multiple terrain features can be used together to serve own mission.	(A) Immediately recognizes and articulates a course of action that will use terrain to own advantage.
(B) Notes terrain features but not significance.	(B) States how individual terrain features could impact own mission (e.g., weather, roads, forests, hills).	(B) Incorporates possible obstacles into plan (e.g., flooding).	(B) Articulates how specific, multiple terrain features will likely be used by enemy and advantages obtained [Visualization]	(B) States action or takes action that will deny identified enemy terrain advantage. [Visualization]
	(C) Seeks information (reports) from subordinates on nature of terrain.	(C) Incorporates terrain features into plan (e.g., uses chokepoints as kill zones; understands accounts for terrain impacts on line of sight.)	(C) Describes *key* aspects of terrain for friendly and/enemy courses of action.	(C) States plan or rationale (or takes action) to deny enemy objective by integrating own use of terrain with likely enemy use.
	(D) Speculates on how terrain could impact evolving mission in a general manner.	(D) Makes a statement about the effects of terrain on assets employed or needed.	(D) Deliberates about best way to use terrain to accomplish the mission.	

1	2	3	4	5
	(E) Speculates about how terrain could impact mission as the situation evolves.	(E) Rejects a route due to terrain conditions.	(E) Makes specific statements about how enemy may leverage terrain (e.g., "the enemy may use that hill for observation").	
		(F) Identifies advantage or disadvantage of a piece of terrain.	(F) Deliberates about how to counteract the way in which the enemy is likely to use the terrain.	
		(G) Identifies terrain features that could cause a problem for specific asset(s).	(G) Describes process by which a judgment should be made about terrain.	
		(H) Describes effects of terrain on course of action or singular actions.	(H) Generates specific ideas about how to use terrain, assets, or other resources to inhibit enemy accomplishing objective.	
		(I) Asks questions or identifies information needed about enemy activity along key terrain.	(I) Describes integrated picture of how terrain will affect asset(s) or mission. [Visualization]	

1	2	3	4	5
		(J) Speculates about how terrain features may offer advantages to enemy.	(J) Identifies information needed about terrain, i.e., features or conditions that must be identified during planning due to implications for mission (e.g., mosques, nature of a road).	
		(K) Makes predictions either generally or specifically about enemy use of terrain.		
		(L) Makes specific statements about how to leverage terrain advantages (e.g., "Can I use the forest to protect my forces from observation?" or "Can I obscure the enemy's vision with smoke?")		
		(M) Describes route features required based on assets used or mission tasks.		

1	2	3	4	5
		(N) Identifies terrain features that are advantageous for enemy.		

Appendix D

Interrater Reliability Protocol

Rater Guidelines and Instructions

Print out the four T-BARS:
Theme 1: Know and Use All Assets
Theme 2: Focus on Mission and Higher's Intent
Theme 3: Model a Thinking Enemy/Populace
Theme 4: Consider Effects of Terrain

Read the first page of each of the four T-BARS in order to get a sense of what the theme is about, and what each of the five levels within the theme is intended to represent with regard to performance and cognition.

Each of the bullets (marked by a letter from 'A' to 'M') within a column describes a behavioral indicator that represents cognitive functioning and domain mental models at that level (1-5) of performance.

Read each data segment. Select the theme to which it corresponds. Then within the theme, select the behavioral descriptor that best describes the data. If you are unable to find a behavioral descriptor that explicitly describes the data, then consider a) looking at another theme, or b) using the general descriptors of each level within the originally selected theme to rate the data. Then record the theme, level, and behavioral descriptor (bullet) you've selected on the coding sheet.

You may use the context provided by surrounding data to code a particular segment if it adds clarity to the participant's response.

If a segment seems unrateable because it lacks the content required to make sense, or if seems to be an aside or otherwise unrelated to the vignette or exercise, then do not rate it. Simply record a dash in that cell on the coding sheet.

If a segment seems to contain elements of multiple themes or multiple levels, then break the segment apart and code each part. (We will count the resultant segments as independent chunks to be coded by all raters.)

As you go through the data, record any issues in the "Notes" column of the coding sheet. For example, if you have difficulty discriminating which of two or three behavioral descriptors best fits a particular data segment; record the options you are having trouble choosing between. If you find any of the behavioral descriptors from the BARS to be confusing, record those issues on a separate sheet of paper.